

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
5 April 2001 (05.04.2001)

PCT

(10) International Publication Number  
**WO 01/23541 A2**

- (51) International Patent Classification<sup>7</sup>: C12N 15/00
- (21) International Application Number: PCT/US00/27065
- (22) International Filing Date: 2 October 2000 (02.10.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/156,953 30 September 1999 (30.09.1999) US
- (71) Applicant (*for all designated States except US*): ALEX-  
ION PHARMACEUTICALS, INC. [US/US]; Suite 360,  
25 Science Park, New Haven, CT 06511 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): FODOR, William,  
L. [US/US]; 236 Wildcat Road, Madison, CT 06443 (US).  
RAMSOONDAR, Jagdeecce, J. [TT/US]; 46 Madison  
Street, Hamilton, NY 13346 (US).
- (74) Agents: WIXON, Henry, N. et al.; Hale and Dorr LLP,  
1455 Pennsylvania Avenue, N.W., Washington, DC 20004  
(US).
- (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ,  
BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK,  
DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,  
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU,  
LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT,  
RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA,  
UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,  
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— *Without international search report and to be republished  
upon receipt of that report.*
- For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*



WO 01/23541 A2

(54) Title: COMPOSITIONS AND METHODS FOR ALTERING GENE EXPRESSION

(57) Abstract: Disclosed are novel compositions and methods useful for modulating expression of a target gene in a cell by insertion of exogenous DNA sequence into the target gene. The compositions and methods of the invention are useful for generation of knockout animals including mammals.

## COMPOSITIONS AND METHODS FOR ALTERING GENE EXPRESSION

### TECHNICAL FIELD

The present invention is directed generally to the biological sciences, including  
5 recombinant genetics and immunology. More particularly, there are described herein non-human  
knockout animals, preferably mammals, in which the expression of one or more genes has been  
altered. Also provided herein are xenograft transplants in which the expression of one or more  
genes has been modulated to prevent or reduce the likelihood of rejection by the transplant  
recipient.

10

### BACKGROUND

The immune response of mammals, including humans, against invading pathogens,  
toxins, and other foreign substances involves many specialized cells that act together.  
Lymphocytes are a class of white blood cells responsible for the specificity of the immune  
system. Two important classes of lymphocytes are T cells and B cells. T cells develop in the  
15 thymus, and are responsible for cell mediated immunity. There are many types of specialized T  
cells, such as for example, helper T cells (which enhance the activity of other types of white  
blood cells), suppressor T cells (which suppress the activity of other white blood cells), and  
cytotoxic T cells (which kill cells). B cells develop in the bone marrow and exert their effect by  
producing and secreting antibodies.

20 A key to the coordinated immune response is complement, which, as described in U.S.  
Patent No. 5,679,345, is involved in the pathogenesis of tissue injury observed in many  
immunologically mediated diseases, such as systemic lupus, erythematosis, rheumatoid arthritis,

and immune-hemolytic anemia. Complement is also involved in rejection of transplanted organ grafts. Complement is responsible for much of the tissue injury in transplantation due to inflammatory conditions resulting from rejection or superimposed by infection, ischemia, and thrombosis of vessels in the graft, as well as tissue injury due to inflammation from similar causes in patients who have not received an organ transplant. In particular, complement attack on cells is central to the rapid onset phase of immune mediated graft rejection (hyperacute rejection), where complement activation and subsequent tissue damage occur within hours.

Graft rejection may occur through a number of different mechanisms, with the time course of rejection being characteristic of the particular mechanism. Early rejection (hyperacute rejection), occurring within minutes or hours of transplantation, involves complement activation by components that are present at the time of the transplant operation. Activation may occur via the classical pathway by preformed antibodies that are reactive with the "foreign" or non-self markers of the graft or via the alternative pathway in response to tissue damage in the graft as a result of, for example, ischemic damage to the organ during storage before transplantation.

Acute rejection occurs days to weeks after transplantation, and is caused by sensitization of the host to the foreign tissue that makes up the graft. Once the host's immune system has identified the transplanted tissue as foreign, all the resources of the immune system are marshaled against the graft, including both specific (antibody and T cell-dependent) responses and non-specific (phagocytic and complement-dependent) responses. Chronic rejection will usually only occur when the graft recipient is immune-suppressed. Then the graft may survive long enough for tissue to undergo changes which ultimately affect survival of the graft. Such changes include hyperplasia and tissue hypertrophy, and endothelial cell damage leading to narrowing of the vascular lumen and potentially impairing the oxygen supply of the graft tissue.

Xenograft rejection of pig tissue is triggered by natural human antibodies that recognize carbohydrate xeno-antigens, such as Gal  $\alpha(1,3)$  galactose, which is expressed on pig endothelial cells that line blood vessels. Weiss, *Science*, 285(20):1221-1222 (August 20, 1999). U.S. Patent No. 5,821,117 describes inhibiting xenotransplant rejection by disrupting the wild type porcine Gal  $\alpha(1,3)$  galactosyl transferase gene with a cloned mutant porcine Gal  $\alpha(1,3)$  galactosyl transferase sequence specifically within an exon of the wild type gene. The resultant mutant gene does not encode a functional galactosyl transferase, with the expected result that rejection of the transplanted xenograft by the patient's immune system is avoided.

In such so called "knockout" mammals, expression of an endogenous gene has been altered (typically, suppressed) through genetic manipulation. Preparation of knockout mammals typically has required introducing into an undifferentiated cell type (termed an embryonic stem cell) a nucleic acid construct to suppress expression of a target gene. This cell is introduced and integrated into a mammalian embryo. The embryo is implanted into a foster mother for the duration of gestation. For example, Pfeffer et al. (*Cell*, 73:457-467 [1993]) describe mice in which the gene encoding the tumor necrosis factor receptor p55 has been disrupted by mutation utilizing homologous recombination. The mice showed a decreased response to tumor necrosis factor signaling. Fung-Leung et al. (*Cell*, 65:443-449 [1991]; *J. Exp. Med.*, 174:1425-1429 [1991]) describe knockout mice lacking expression of the gene encoding CD8. These mice were found to have a decreased level of cytotoxic T cell response to various antigens and to certain viral pathogens such as lymphocytic choriomeningitis virus.

Typical prior methods, however, describe manipulation of an exon region of the target gene. There is thus a need in the art for new and improved methods for modulating gene

expression in animals including mammals, particularly for overcoming xenograft transplant rejection. It is to these, as well as other, important ends that the following is addressed.

### SUMMARY

It has been discovered that the expression of a particular gene in an animal may be modulated by introducing into the genomic DNA of the animal a new DNA sequence that results in the disruption of at least some portion of the DNA sequence of the gene to be modulated. The methods described herein are of general utility for altering gene expression in animals including mammals. In contrast to prior methods, it has surprisingly been found that gene expression may be suppressed in part or in total by inserting new DNA sequence into the intron of the target genomic DNA.

The versatility of the methods described herein for generating "knockout" animals is illustrated by the following general description of a preferred embodiment, including the examples. It is to be understood that while the remaining discussion is directed largely to the utility of Gal  $\alpha(1,3)$ galactosyl transferase knockout pigs, the utility of the methods described herein is not limited to solely this protein. Rather, the following discussion is provided merely for exemplification of their versatility and preferred use.

### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows the nucleotide sequence of introns of the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene from within intron 3 to the end of intron 8. Dashes indicate nucleotides within an exon region. Thus, nucleotide sequence numbering represents the number of bases in the entire porcine Gal  $\alpha(1,3)$  galactosyl transferase gene relative to nucleotide position 1 of the insert isolated from the lambda-2 phage clone.

FIG. 2 shows a schematic representation of the gene targeting vector used for inactivation of the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene (see Example 1). This vector is designed to contain a sequence with homology to the 5' region of intron 3 of the Gal  $\alpha(1,3)$  galactosyl transferase gene, a promoterless neomycin phosphotransferase gene engineered to contain multiple stop codons (engineered exon), an engineered splice acceptor site, the 5' region of intron 4 sequence for splicing the engineered exon to the downstream exon 4, and a sequence with homology to the 3' region of intron 3 to aid with annealing to the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene. Arrows indicate location of primers used for PCR.

FIG. 3 shows the nucleotide sequence of the gene targeting vector used for inactivation of the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene (see Example 1). This vector is designed to contain (A.) a sequence with homology to the 5' region of intron 3 of the Gal  $\alpha(1,3)$  galactosyl transferase gene, (B.) an intron 4 splice acceptor sequence, (C.) a promoterless neomycin phosphotransferase gene engineered to contain multiple stop codons (engineered exon), (D.) an intron 4 splice donor signal sequence, and (E.) a 3' intron 3 sequence to aid with annealing to the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene. All underlined sequences correspond to restriction sites in the primer sequences. Bold type indicates primer regions used for PCR. Normal type indicates PCR fragment sequences.

FIG. 4 shows the nucleotide sequence for the neomycin phosphotransferase gene (the neomycin resistance gene). Bold type indicates the location of gene start and stop codons. The underlined sequence corresponds to primer sequences. Nucleotides which are capitalized are within the coding region of this gene.

FIG. 5 shows the nucleotide sequence of the puromycin/bovine growth hormone poly A. The underlined sequences correspond to the puromycin gene start codon.

FIG. 6 shows a schematic representation of the gene targeting vector used for inactivation of the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene (see Example 2). This vector is designed to contain a sequence with homology to the Gal  $\alpha(1,3)$  galactosyl transferase gene 3' intron 3 sequence including the 3' intron splice acceptor sequence, a Kozak consensus sequence, a promoterless puromycin gene engineered to contain a bovine growth hormone poly A sequence (engineered exon), and a sequence with 5' intron 4 sequence homology including the 5' intron splice donor sequence. Arrows indicate location of primers used for PCR.

FIG. 7 shows the nucleotide sequence of the gene targeting vector shown schematically in FIG. 6 (see Example 2). The underlined sequences correspond to the primer sequences used. Bold type indicates the intron regions used for homology. The AG and GT splice consensus sequences at the 3' end of intron 3 and the 5' end of intron 4 are in upper case.

FIG. 8 shows the nucleotide sequence of the ricin A toxin gene. Nucleotides which are capitalized are within the coding region of this gene.

FIG. 9 shows a schematic representation of the collision construct used for inactivation of the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene (see Example 3). This vector is designed to contain a sequence with homology to the Gal  $\alpha(1,3)$  galactosyl transferase gene 3' intron 3 sequence including the 3' intron splice acceptor sequence, a reverse orientation puromycin gene engineered to contain a bovine growth hormone poly A sequence under the control of a phosphoglycerate kinase (PGK) promoter, and a sequence with 5' intron 4 sequence homology including the 5' intron splice donor sequence, and a ricin A toxin gene under the control of a cytomegalovirus (CMV) promoter and containing a SV40 poly A sequence located outside the regions of homology. Arrows indicate location of primers used for PCR.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

Technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art, unless otherwise defined herein. Although any methods and materials similar or equivalent to those described herein may be used in the practice or testing of the described methods, preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications which might be used in connection herewith.

As used herein and in the appended claims, the singular forms "a," "an," and "the" are intended to include the plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" is intended to include a plurality of such host cells, reference to "an antibody" is intended as a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth. It is to be understood that the appended claims are not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, which those of skill will appreciate may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which is to be limited only by the appended claims.

The term "knockout" refers to the modulation of the expression of at least a portion of a protein encoded by the target gene. The term "knockout construct" refers to a nucleic acid sequence that is designed to modulate a protein encoded by endogenous DNA sequences in a cell. The nucleic acid sequence used as the knockout construct is typically comprised of DNA from some portion of the gene or genes (including, but not limited to, the exon sequence, intron



sequence, and/or promoter sequence) to be modulated and a sequence marker used to disrupt and select for the presence of the knockout construct in the cell. The nucleic acid sequence of the knockout construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to prevent or interrupt protein expression from the native gene. Such  
5 insertion usually occurs by homologous recombination (i.e., regions of the knockout construct that are homologous to endogenous DNA sequences hybridize to each other when the knockout construct is introduced into the cell and recombines so that the knockout construct is incorporated into the corresponding position of the endogenous DNA).

The knockout construct nucleic acid sequence may comprise a full or partial sequence of  
10 one or more exons and/or introns of the gene to be modulated, a full or partial promoter sequence of the gene to be modulated, or combinations thereof. In one embodiment of the invention, the nucleic acid sequence of the knockout construct comprises a first nucleic acid sequence region homologous to a first nucleic acid sequence region of the gene to be modulated, and a second nucleic acid sequence region homologous to a second nucleic acid sequence region of the gene to  
15 be modulated. The orientation of the knockout construct should be such that the first nucleic acid sequence is upstream of the second nucleic acid sequence and the sequence marker should be therebetween.

A suitable nucleic acid sequence region(s) should be selected so that there is homology between knockout construct sequence(s) and the gene of interest. Preferably, the knockout  
20 construct sequences are isogenic sequences with respect to the target sequences. The nucleic acid sequence region of the knockout construct may correlate to any region of the gene provided that it is homologous to the gene. A nucleic acid sequence is considered to be "homologous" if it is at least about 90% identical, preferably at least about 95% identical, or most preferably, about

100% identical to the nucleic acid sequence. Furthermore, the 5' and 3' nucleic acid sequences flanking the selectable marker should be sufficiently large to provide complementary sequence for hybridization when the knockout construct is introduced into the genomic DNA of the target cell. For example, homologous nucleic acid sequences flanking the selectable marker gene  
5 should be at least about 500 bp, preferably, at least about 1 kilobase (kb), more preferably about 2-4 kb, and most preferably about 3-4 kb in length. In a preferred embodiment, both of the homologous nucleic acid sequences flanking the selectable marker gene of the construct should be at least about 500 bp, preferably, at least about 1 kb, more preferably about 2-4 kb, and most preferably about 3-4 kb in length.

10 Another suitable DNA sequence includes cDNA sequence provided the cDNA is sufficiently large. Each of the flanking nucleic acid sequences used to make the construct is preferably homologous to one or more exon and/or intron regions, and/or a promoter region. Each of these sequences is different from the other, but may be homologous to regions within the same exon and/or intron. Alternatively, these sequences may be homologous to regions within  
15 different exons and/or introns of the gene. Preferably, the two flanking nucleic acid sequences of the knockout construct are homologous to two sequence regions of the same or different introns of the gene of interest. In addition, it is preferred that isogenic DNA is used to make the knockout construct of the present invention. Thus, the nucleic acid sequences obtained to make the knockout construct are preferably obtained from the same cell line as that being used as the  
20 target cell.

In accordance with the present invention, the integration of the knockout construct nucleic acid sequence into at least one gene of interest results in the modulation of the expression of the gene product. "Modulating" the expression of a gene includes suppressing the expression

of the gene, disrupting the expression of the gene, eliminating the expression of the gene, altering the expression of the gene, or decreasing the expression of the gene relative to expression of the wild-type gene. Preferably, the integrated knockout construct results in reduced protein function relative to native protein function. Most preferably, the integrated knockout construct results in the production of a non-functional protein. Complete or absolute non-functionality of the protein is not required.

The phrases "disruption of the gene" and "gene disruption" refer to insertion of a nucleic acid sequence into at least one region of the native DNA sequence (usually one or more exons or one or more introns) and/or the promoter region of a gene so as to modulate expression of that gene in the cell as compared to the wild-type or naturally occurring sequence of the gene. By way of example, a nucleic acid construct may be prepared containing a DNA sequence encoding an antibiotic resistance gene which is inserted between the DNA sequence complementary to the target gene DNA sequence (promoter and/or coding region) to be disrupted. When this nucleic acid construct is then transfected into a cell, the construct will integrate into the genomic DNA either randomly or into the target gene by homologous recombination. It has been found that selection for drug resistant cells in the population of transfectants enhance the probability of obtaining a homologous gene knockout. Thus, many progeny of the cell will no longer express the gene, or will express it at a decreased level, as the DNA is now disrupted by the antibiotic resistance gene.

In some instances, such as, for example, where the methods described herein are used to produce cells, tissues or organs suitable for xenotransplant into humans, it may not be necessary to completely eliminate the production of functional protein. Rather, it will be satisfactory to reduce the production of functional protein only to a level that will, in conjunction with other

therapeutic regimens, prevent or reduce the patient's immune response and the likelihood of rejection. Thus, for example, a knockout achieved according to the methods described herein may preferably reduce the biological activity of the polypeptide normally encoded therefrom by at least about 70%, preferably at least about 80%, relative to the unmutated gene.

5           The knockout construct may be inserted into any suitable target cell for integration into its genomic DNA that may be maintained in culture. Suitable cells include, but are not limited to, fibroblast, epithelial cell, endothelial cell, transgenic embryonic fibroblast, embryonic stem cell, and primordial germ cell. In one embodiment, the knockout construct is inserted into an embryonic stem cell (ES cell) and is integrated into the ES cell genomic DNA. ES cells  
10           comprising the integrated knockout construct are then injected into, and integrate with, a developing mouse embryo. In another embodiment, the knockout construct is inserted into a nuclear transfer donor cell. Suitable nuclear transfer donor cells include fibroblasts, epithelial cells, and cumulous cells. In this embodiment, the knockout construct is inserted into the nuclear transfer donor cell, and the donor cells comprising the knockout construct are fused with an  
15           enucleated oocyte. The resultant fused oocyte is then transferred to a surrogate female. Furthermore, where the target cell is intended to be used to produce a knockout mammal, it is preferred that the target cell be derived from the same species as the knockout mammal to be generated. Thus, for example, pig embryonic stem cells or pig fibroblasts will usually be used for generation of knockout pigs.

20           The nucleic acid sequence of the knockout construct may be integrated into the genomic DNA of the host cell using any suitable method. In one preferred embodiment, integration is achieved by the process of homologous recombination. Homologous recombination has been described previously, for example, in Kucheripati *et al.* (1984) Proc. Natl. Acad. Sci. USA

81:3153-3157; Kucherlapati *et al.* (1985) *Mol. Cell. Bio.* 5:714-720; Smithies *et al.* (1985) *Nature* 317:230- 234; Wake *et al.* (1985) *Mol. Cell. Bio.* 8:2080-2089; Ayares *et al.* (1985) *Genetics* 111:375-388; Ayares *et al.* (1986) *Mol. Cell. Bio.* 7:1656- 1662; Song *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:6820-6824; Thomas *et al.* (1986) *Cell* 44:419-428; Thomas and  
5 Capecchi (1987) *Cell* 51:503- 512; Nandi *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:3845-3849; and Mansour *et al.* (1988) *Nature* 336:348-352, which are herein incorporated by reference. Furthermore, various aspects of using homologous recombination to create specific genetic mutations in embryonic stem cells and to transfer these mutations to the germline have been described. (Evans and Kaufman (1981) *Nature* 294:154-146; Doetschman *et al.*, (1987)  
10 *Nature* 330:576-578; Thomas and Capecchi (1987) *Cell* 51:503-512; Thompson *et al.* (1989) *Cell* 56:316-321.) In homologous recombination, DNA fragments between two DNA molecules are exchanged during crossover at the site of the homologous nucleic acid sequences. Thus, crossover would occur between the knockout construct and eukaryotic gene at the site of homology within the 5' region of the first nucleic acid sequence of the construct (homologous to  
15 the first nucleic acid region of the gene of interest). A second crossover event would occur in the 3' region of the construct homologous to the second nucleic acid region of the gene of interest. As a result, the sequence information between these two regions of the knockout construct would be inserted into the gene of interest in the host cell's genomic DNA.

The methods described herein may be used to produce a mammal in which one, two, or  
20 more genes have been knocked out. Such mammals may be generated by repeating the procedures set forth herein for generating each knockout construct, or by breeding two mammals, each with a different single gene knocked out, to each other, and screening for those with the double knockout genotype.

The term "marker sequence" or "selectable marker" refers to a nucleic acid sequence that is used as part of the knockout construct to modulate the expression of the gene of interest, and as a means to identify those cells that have incorporated the knockout construct into the genome. The selectable marker may be any sequence that serves these purposes. For example, the

5 selectable marker may encode a protein that confers a detectable trait on the cell, such as an antibiotic resistance gene, or an assayable enzyme not typically found in the target cell. The selectable marker gene may be any nucleic acid sequence that is detectable and/or assayable, which is used to recover transformed cell lines. One having skill in the art will be capable of determining suitable selectable markers for use in the present invention. For example, suitable

10 selectable markers include, but are not limited to,  $\beta$ -lactamase (ampicillin resistance), kanamycin resistance, gentamicin resistance, puromycin-N-acetyl-transferase, hygromycin b-phosphotransferase, thymidine kinase, and tryptophan synthetase. For example, the herpes simplex virus thymidine kinase (tk) (Wigler, M. et al. (1977) Cell 11:223-32) or adenine phosphoribosyltransferase (aprt) (Lowy, I. et al. (1980) Cell 22:817-23) genes, which may be

15 employed in tk or aprt cells, respectively, may be used as the selectable marker. Also, antimetabolite, antibiotic or herbicide resistance may be used as the basis for selection; for example, dihydrofolate reductase (dhfr), which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); neomycin phosphotransferase (npt), which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol.

20 Biol. 150:1-14). Additional selectable genes have been described and include, for example, tryptophan synthetase (trpB), which allows cells to utilize indole in place of tryptophan, or histidinol dehydrogenase (hisD), which allows cells to utilize histidinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047- 51). Recently, the

use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase (GUS), and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) Methods Mol. Biol.

5 55:121-131). In the present invention, it is preferred that the selectable marker gene is an antibiotic resistance gene, such as the neomycin resistance gene or puromycin resistance gene.

Moreover, when the selectable marker encodes a protein, it may also contain a promoter that regulates its expression, or require expression from an endogenous promoter, preferably the target gene promoter. Thus, the selectable marker gene may be operably linked to its own  
10 promoter or be promoterless. The selectable marker gene may be inserted into the knockout construct without its own promoter attached as it may be transcribed using the promoter of the gene to be suppressed. In addition, the marker gene may have a polyA signal sequence attached to the 3' end of the gene, which serves to terminate transcription of the gene and process the transcript with the addition of adenine residues at the 3' end to stabilize the mRNA.

15 In one embodiment a target gene (*e.g.*, Gal  $\alpha(1,3)$  galactosyl transferase) is modulated by insertion of an engineered exon or active gene within an intron of the target gene. In this embodiment, the target gene is prevented from being translated by insertion of an in-frame, promoterless engineered exon (*e.g.*, an antibiotic resistance gene) that contains multiple stop codons within an intron of the target gene. Using this 'promoter-trap' strategy, the engineered exon is spliced  
20 in frame upstream of the exon comprising the start codon. This results in the expression of the drug resistance gene prior to the gene of interest and concomitantly inhibits expression of the target gene due to the presence of multiple stop codons downstream of the drug resistance gene. As described

herein, any gene that confers survival of the targeted cells under appropriate selection conditions may be used as the engineered exon.

Using the 'promoter trap' strategy, a gene targeting construct is designed which contains a sequence with homology to an intron sequence of the target gene (*e.g.*, the intron 3 sequence of the Gal  $\alpha(1,3)$  galactosyl transferase gene), a downstream intron splice acceptor signal sequence  
5 comprising the AG dinucleotide splice acceptor site (*e.g.*, the intron 4 splice acceptor signal sequence of the Gal  $\alpha(1,3)$  galactosyl transferase gene), a promoterless selectable marker engineered exon (*e.g.*, drug resistance gene) engineered to contain multiple stop codons, the intron splice donor signal sequence comprising the GT dinucleotide splice donor site (*e.g.*, the intron 4 splice donor sequence of  
10 the Gal  $\alpha(1,3)$  galactosyl transferase gene) for splicing the engineered exon to the immediate downstream exon (*e.g.*, exon 4 of the Gal  $\alpha(1,3)$  galactosyl transferase gene), and additional sequence with homology to the intron sequence of the target gene (*e.g.*, intron 3 sequence homology of the Gal  $\alpha(1,3)$  galactosyl transferase gene ) to aid with annealing to the target gene. It will be appreciated that the method may be used to target any intron within target gene of interest.

15 In another embodiment, the 'promoter trap' strategy is used to modulate target gene expression by replacing an endogenous exon with an in-frame, promoterless engineered exon (*e.g.*, an antibiotic resistance gene). The engineered exon is spliced in frame and results in the expression of the drug resistance gene and concomitant inhibited expression of the full-length target gene.

20 This 'promoter trap' gene targeting construct may be designed to contain a sequence with homology to the target gene 3' intron sequence upstream of the start codon (*e.g.*, the Gal  $\alpha(1,3)$  galactosyl transferase gene 3' intron 3 sequence), the upstream intron splice acceptor sequence comprising the AG dinucleotide splice acceptor site (*e.g.*, the intron 3 splice acceptor sequence), a



Kozak consensus sequence, a promoterless selectable marker gene containing *e.g.*, a poly A termination sequence (the engineered exon), a splice donor sequence comprising the GT dinucleotide splice donor site from an intron region downstream of the start codon (*e.g.*, the 5' intron 4 splice donor sequence), and a sequence with 5' sequence homology to the downstream intron (*e.g.*, 5' intron 4). It will be appreciated that the method may be used to target any exon within the Gal  $\alpha(1,3)$  galactosyl transferase gene or any other gene of interest. A representative construct useful for targeting the pig Gal  $\alpha(1,3)$  galactosyl transferase gene was deposited with the American Type Culture Collection (ATCC) on 28 September 2000 with accession number \_\_\_\_\_ and is described herein in Example 2.

In yet another embodiment, the selectable marker may be inserted into the knockout construct in a reverse orientation to the targeted gene. In this embodiment, a strong promoter is used with the selectable marker all in the reverse orientation, which drives transcription in the reverse direction and therefore, modulates the expression of the targeted gene. The target gene is modulated using a "collision construct" to insert an active gene in place of an exon and at least part of the flanking introns, including the splice donor and splice acceptor sites. The inserted gene, such as a selectable marker gene, is under the control of a highly active promoter such as the phosphoglycerate kinase I (PGK) gene promoter, such that transcription of this gene causes the termination of transcription of the endogenous gene (Rosario *et al.*, (1996) Nat. Biotech. 14:1592-1596). The selectable marker gene is further engineered to contain a transcription termination sequence. Insertion of the engineered gene may be made to replace any exon, within any intron, or portions thereof to result in a truncated transcript which modulates the expression of a functional target gene product. It will be appreciated that this method may be used to target any intron or exon of interest of the target gene. Positive selection for transfected cells in which the construct has been integrated may be accomplished via expression of the selectable marker gene. As described herein, it will be appreciated

that any selection marker gene that confers survival of the targeted cells under appropriate selection conditions may be driven by the strong PGK promoter. Additionally, a toxin gene (*e.g.*, Ricin A toxin) is preferably engineered into the collision construct inserted to eliminate random integration events. A representative collision construct useful for targeting the pig Gal  $\alpha(1,3)$  galactosyl transferase gene has been deposited with the ATCC on 28 September 2000 with accession number \_\_\_\_\_ and is described herein in Example 3.

The integrated selectable marker nucleic acid in the cell is capable of modulating the expression of the gene of interest. Expression of the selectable marker allows for selection of the cells which comprise the integrated sequence. Modulation of the expression of the gene of interest is accomplished by disruption of the endogenous gene by an engineered exon in forward or reverse orientation with the endogenous gene.

The term "animal," as used herein, is intended to include any multicellular eukaryotic organism, preferred among which are mammals. When used in the context of a xenograft donor, the term "mammal" preferably includes, but is not limited to, pigs, sheep, goats, cows, deer, rabbits, hamsters, rats, mice, horses, cats, dogs, and the like. Preferably, humans are excluded.

The term "progeny" refers to any and all future generations derived or descending from a particular mammal, *i.e.*, a mammal containing a knockout construct inserted into its genomic DNA. Thus, progeny of any successive generation are included herein such that the progeny, the F1, F2, F3 generations, and so on indefinitely, are included in this definition.

The terms "immunomodulate" and "immunomodulation" refer to changes in the level of activity of any components of the immune system as compared to the average activity of that component for a particular species. Thus, as used herein, immunomodulation refers to an increase or a decrease in activity. Preferably, in accordance with the present invention, the

integration of the selectable marker into the gene of interest results in a decreased immune response, when the host cell is introduced to a patient. Immunomodulation may be detected by assaying the level of antibody reactivity, complement activity, B cells, any or all types of T cells, antigen presenting cells, and any other cells believed to be involved in immune function.

5 Additionally or alternatively, immunomodulation may be detected by evaluating the level of expression of particular genes believed to have a role in the immune system, the level of particular compounds such as cytokines (interleukins and the like) or other molecules that have a role in the immune system, and/or the level of particular enzymes, proteins, and the like that are involved in immune system functioning.

10 The target gene to be knocked out may be any gene, provided that at least some sequence information on the DNA to be disrupted is available to use in the preparation of both the knockout construct and the screening probes. It is not necessary that the entire genomic sequence of the target gene be known in order to use the methods described herein.

The target gene to be knocked out preferably will be a gene that is expressed in mature  
15 and/or immature T cells and/or B cells. It is a further preference that the target gene is expressed in target antigen presenting cell, target endothelial cell, target neuronal cell, or any target cell that may be attacked by the humoral or cellular immune system of the recipient. The target gene is further preferably involved, either directly or indirectly, in the activation pathway during inflammation or immunosuppression responses by the immune system, and does not result in  
20 lethality when knocked out. In accordance with the present invention, expression of target genes may advantageously be altered according to the methods described herein to produce xenotransplant cells, tissues and organs for use in humans, in order to reduce or prevent immune response and rejection by the patient.

Thus, in accordance with the present invention, any gene may be used provided that it can undergo homologous recombination and the expression of which may be modulated by insertion of the knockout construct of the present invention. Suitable genes include, but are not limited to, B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2 or VCAM-1, CD28, CD80, CD86, CD154, major histocompatibility complex class I,  $\beta$ -2-microglobulin, invariant chain (Ii), caspase-1, caspase-3, and Gal  $\alpha$ (1,3) galactosyl transferase gene. This list is not intended to be exhaustive. One having ordinary skill in the art would be capable of ascertaining suitable genes to be modulated. Preferably, the gene is implicated in the immunoresponse system of a patient. More preferably, the target gene is a porcine target gene selected from the group consisting of CD 80, CD 86, B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2 or VCAM-1. A presently preferred porcine target gene is the Gal  $\alpha$ (1,3) galactosyl transferase gene.

The DNA sequence to be used to knock out a selected gene may be obtained using methods well known in the art such as those described by Sambrook et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)).

Such methods include, for example, screening a genomic library with a cDNA probe encoding at least a portion of the same gene in order to obtain at least a portion of the genomic sequence.

Alternatively, if a cDNA sequence is to be used in a knockout construct, the cDNA may be obtained by screening a cDNA library with oligonucleotide probes or antibodies (where the library is cloned into an expression vector). If a promoter sequence is to be used in the knockout construct, synthetic DNA probes may be designed for screening a genomic library containing the promoter sequence. Another method for obtaining the DNA to be used in the knockout construct is to manufacture the DNA sequence synthetically, using a DNA synthesizer.

In another embodiment, porcine genomic DNA encoding the Gal  $\alpha(1,3)$  galactosyl transferase gene is isolated from a lambda phage clone library. A pig genomic library is screened using a cDNA corresponding to exon 4 of the Gal  $\alpha(1,3)$  galactosyl transferase gene. Phage are screened and unique clones, which contain exon 4 sequences are isolated using standard library screening methods. (Sambrook et al.) Clones obtained by this procedure contain inserts 15-40 kb in length. These clones, were designated pgGT, lambda 1, lambda 2, lambda 4-1 and lambda 8-2. Five vectors comprising unique, overlapping nucleotide sequences which span the entire the pig Gal  $\alpha(1,3)$  galactosyl transferase gene from within intron 3 through intron 8 have been deposited with the ATCC: (1) a 1.6 kb insert within intron 3 of the extreme 5' end of the 18.275 kb lambda-2 phage clone, (2) a 6.7 kb HindIII fragment spanning intron 3 to intron 4 of the 18.275 kb lambda-2 phage clone, (3) a 4 kb HindIII fragment following the 6.7 kb fragment 2 of the 18.275 kb lambda-2 phage clone, (4) a 6 kb HindIII-SalI fragment at the 3' most portion of the 18.275 kb lambda-2 phage clone, and (5) a 13 kb fragment of the lambda-2 phage clone spanning exon 7 to exon 9. These five vectors were deposited with ATCC on 29 September 2000 with accession numbers \_\_\_\_\_, respectively. Subclones of the various inserts were used to generate the claimed intron sequences from within intron 3 to intron 8 as provided in Figure 1. These sequences may be used to determine regions of sequence homology in design of targeting constructs for modulation of the pig Gal  $\alpha(1,3)$  galactosyl transferase gene.

The DNA sequence encoding the knockout construct is preferably generated in sufficient quantity for genetic manipulation and insertion the target cell. Amplification may be accomplished by known methods, such as by placing the sequence into a suitable vector and transforming bacterial or other cells that may rapidly amplify the vector, by PCR amplification, or by synthesis with a DNA synthesizer.

The DNA sequence to be used in producing the knockout construct is digested with a restriction enzyme selected to cut at a location(s) such that a new DNA sequence encoding a selectable marker gene may be inserted in the proper position within this DNA sequence. The proper position for a selectable marker gene insertion is that which will serve to modulate

5 expression of the native gene. This position will depend on various factors such as the restriction sites in the sequence to be cut, and whether, for example, an intron sequence, an exon sequence or a promoter sequence is (are) to be modulated. In other words, the precise location of insertion of the selectable marker into the DNA sequence is that which will result in the modulation of promoter function or of synthesis of the native exon. For example, the knockout construct may

10 be engineered to insert the selectable marker entirely within a single intron of the target gene. In this manner, the first nucleic acid sequence would comprise a region of the selected intron upstream from the second nucleic acid sequence and the second nucleic acid sequence would be selected comprising a region of the selected intron located downstream of the first nucleic acid sequence. The selectable marker would be introduced between the first and second nucleic acid

15 sequences. When the construct is then introduced to the cell, the construct nucleic acid sequence is integrated into the target gene and the selectable marker is inserted entirely within the targeted intron.

Similarly, the construct may be engineered to insert the selectable marker within any desired and suitable region of the gene provided that expression is modulated. For example, the

20 construct may be engineered to insert the selectable marker between two adjacent introns and thereby completely remove an endogenous exon of the target gene, to span over a region comprising at least a portion of an intron and at least a portion of an adjacent intron of the targeted gene, to span over a region comprising at least a portion the promoter for the targeted

gene to an adjacent intron, to span over a region encompassing more than one targeted gene, and combinations thereof.

After the genomic DNA sequence has been digested with the appropriate restriction enzymes, the selectable marker gene is ligated into the genomic DNA sequence using methods well known to the skilled artisan and described in Sambrook *et al.*, *supra*. The ends of the DNA  
5 fragments to be ligated must be compatible; this is achieved by either cutting all fragments with enzymes that generate compatible ends, or by blunting the ends prior to ligation. Blunting is done using methods well known in the art, such as for example by the use of Klenow fragment (DNA polymerase I) to fill in sticky ends.

10 The ligated knockout construct may be introduced directly into the target cell, or it may first be placed into a suitable vector for amplification prior to insertion. Preferred vectors are those that are rapidly amplified in bacterial cells such as the pBluescript II SK vector (Stratagene, San Diego, Calif.) or pGEM7 (Promega Corp., Madison, Wis.).

In another embodiment of the invention, embryonic stem (ES) cells are used as the target  
15 cell for their ability to integrate into and become part of the germ line of a developing embryo so as to create germ line transmission of the knockout construct. Thus, any ES cell line that is believed to have this capability is suitable for use herein. For example, one mouse strain that is has been used for production of ES cells is the 129J strain. The cells are cultured and prepared for DNA insertion using methods well known to the skilled artisan such as those set forth by  
20 Robertson (Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E. J. Robertson, ed. IRL Press, Washington, DC (1987)), Bradley et al. (*Current Topics in Devel. Biol.*, 20:357-371 (1986)), and Hogan et al. (Manipulating the Mouse Embryo: A Laboratory Manual, Cold

Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1986)), all of which are incorporated by reference herein.

Insertion of the knockout construct into the target cells may be accomplished using a variety of transfection methods well-known in the art. For example, suitable transfection  
5 methods include electroporation, microinjection, and calcium phosphate treatment (see Lovell-Badge, in Robertson, ed., *supra*). A preferred method of transfection is electroporation. If the cells are to be electroporated, the targeted cells and knockout construct DNA are exposed to an electric pulse using an electroporation machine and following the manufacturer's guidelines for use. After electroporation, the cells are allowed to recover under suitable incubation conditions.  
10 The cells are then screened for the presence of the knockout construct.

Each knockout construct DNA to be introduced into the cell must first be linearized if the knockout construct has been inserted into a vector. Linearization is accomplished by digesting the DNA with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the knockout construct sequence.

15 For introduction of the DNA sequence, the knockout construct DNA is added to the target cells under appropriate conditions for the insertion method chosen. Where more than one construct is to be introduced into the target cell, DNA encoding each construct may be introduced simultaneously or one at a time.

Screening may be done using methods known in the art or combinations thereof. Where  
20 the selectable marker gene is an antibiotic resistance gene, the cells are cultured in the presence of an otherwise lethal concentration of the antibiotic. Those cells that survive have presumably integrated the knockout construct. If the selectable marker gene is other than an antibiotic resistance gene, the genomic DNA of the target cell may be extracted from the cells using



standard methods such as those described by Sambrook *et al.*, *supra*. The DNA may then be probed on a Southern blot with a probe or probes designed to hybridize only to the selectable marker sequence. If the selectable marker gene is a gene that encodes an enzyme whose activity may be detected (e.g., beta-galactosidase), the enzyme substrate may be added to the cells under  
5 suitable conditions, and an appropriate assay for enzymatic activity may be conducted. In addition, the genomic DNA may be amplified by polymerase chain reaction (PCR) with probes specifically designed to amplify DNA fragments of a particular size and sequence (i.e., only those cells containing the knockout construct in the proper position will generate DNA fragments of the proper size). PCR may be used in detecting the presence of homologous recombination  
10 (Kim and Smithies, (1988) Nucleic Acid Res. 16:8887-8903; Joyner et al (1989) Nature 338:153-156). Primers may be used which are complementary to a sequence within the construct and complementary to a sequence outside the construct and at the target locus. In this way, one may only obtain DNA duplexes having both of the primers present in the complementary chains in which homologous recombination has occurred. By demonstrating the  
15 presence of the primer sequences or the expected size sequence, the occurrence of homologous recombination is supported.

Upstream and/or downstream from the target gene knockout construct may be inserted a gene which provides for identification of whether a double crossover has occurred. For this purpose, any suitable marker may be used for as described herein. Preferably, the selectable  
20 marker used to identify double crossovers is different than the selectable marker used to identify the integration of the target gene knockout construct. In one preferred embodiment, the herpes simplex virus thymidine kinase gene is employed, since the presence of the thymidine kinase gene may be detected by the use of nucleoside analogs, such as Acyclovir or Gancyclovir, for

their cytotoxic effects on cells that contain a functional HSV-tk gene. The absence of sensitivity to these nucleoside analogs indicates the absence of the thymidine kinase gene and, therefore, where homologous recombination has occurred, a double crossover event has also occurred.

The knockout construct may be integrated into several locations in the target cell genome, and may integrate into a different location in each cell's genome, due to the occurrence of random insertion events. Notwithstanding random multiple integration sites, the desired location of the insertion is in a complementary position to the DNA sequence to be knocked out. It has been found that less than about 1-5% of the targeted cells that take up the knockout construct will actually integrate the knockout construct in the desired location. Identification of those cells with proper integration of the knockout construct is described herein.

In one embodiment of the present invention, suitably transfected target cells containing the knockout construct in its proper location are inserted into an embryo. Insertion may be accomplished in any suitable method known in the art. Preferably, the cells are introduced into the embryo by microinjection. Most preferably, the cells are ES cells for injection into mouse embryos. For microinjection, about 10-30 cells are collected into a micropipet and injected into embryos that are at the proper stage of development to integrate the transfected cell into the developing embryo. The suitable stage of development for injecting into the embryo is prior to the formation of the germinal layer of the developing embryo as one having ordinary skill in the art may readily determine. Preferably, the embryo is in the early blastocyst stage. By way of example, mice embryos may be introduced to the transfected cells in about 3.5 days. The embryos are obtained by perfusing the uterus of pregnant females by methods known to the skilled artisan (e.g., Bradley (in Robertson, ed., *supra*)). Preferably, the embryos are male.

After the transfected target cell having proper integration of the target gene has been introduced into the embryo, the embryo is implanted into the uterus of a pseudopregnant foster mother. While any foster mother may be used, selection of the foster mother is based upon its ability to breed and reproduce well, and to care for its young. Such foster mothers are typically prepared by mating with vasectomized males of the same species. The stage of the pseudopregnant foster mother is important for successful implantation, and is species dependent. For mice, this stage is about 2-3 days pseudopregnant.

In another embodiment, the suitable transfected target cells are nuclear transfer donor cells. Nuclear transfer donor cells may be virtually any somatic cell type and include fibroblasts, epithelial cells, cumulus cells, *etc.* Nuclear transfer donor cells are cultured *in vitro* and targeted using the constructs and techniques described herein via homologous recombination. Cells are grown in the appropriate medium to allow for selection of cells comprising the having properly integrated the knockout construct. PCR may also be done for confirmation of correctly targeted integration. Thereafter, an unfertilized oocyte of an animal is enucleated using known methods. The enucleated unfertilized oocyte is then fused to the selected knockout nuclear transfer donor cell. Fusion may be conducted by electrical stimulation, chemical stimulation, insertion by injection, or other known methods. The fused product is then cultured, assessed for viability and transferred to a surrogate recipient female. For reference and methods, see *e.g.*, Campbell *et al.* (1996) *Nature* 380:64; Wilmut *et al.* (1997) *Nature* 385:810; WO00/25578; WO97/07669; WO99/36510; WO00/42174; WO99/53751; WO99/45100, which are incorporated herein by reference.

Offspring or progeny that are born to the foster mother or surrogate recipient female are screened (*e.g.*, by PCR) for genomic DNA comprising the knockout construct. This step is

particularly important for selecting for progeny of foster mothers that carried embryos in which the transfected target cell was injected. On the other hand, the progeny of surrogate recipient females that carried the transfected target cell fused with the enucleated unfertilized oocyte will typically have the knockout construct inserted into its genome.

5           Any suitable selection method may be used. For example, if a coat color selection strategy has been used, the offspring may be screened for a coat color indicative of proper integration of the targeted gene into the offspring. Other methods include obtaining DNA from the offspring and screening for the presence of the knockout construct using Southern blots and/or PCR as described herein. Other means of identifying and characterizing the knockout  
10   offspring include the use of Northern blots and Western blots. For example, Northern blots may be used to probe the mRNA for the presence or absence of transcripts encoding either the gene knocked out, the marker gene, or both. In addition, Western blots may be used to assess the level of expression of the gene knocked out in various tissues of these offspring by probing the Western blot with an antibody against the protein encoded by the gene knocked out, or an  
15   antibody against the marker gene product, where this gene is expressed. *In situ* analysis (such as fixing the cells and labeling with antibody) and/or fluorescence activated cell sorting (FACS) analysis of various cells from the offspring may be conducted using suitable antibodies to look for the presence or absence of the knockout construct gene product.

          Offspring that appear to contain the integrated knockout construct in its genome may then  
20   be out-crossed to generate multiple offspring if they are believed to carry the knockout construct in their germ line to generate F1 offspring heterozygous for the knockout construct. F1's will then be crossed to generate homozygous knockout animals.

The heterozygotes may then be crossed with each other to generate homozygous knockout offspring. Homozygotes may be identified by any screening method as described herein. For example, the homozygotes may be identified by Southern blotting of equivalent amounts of genomic DNA from the host animal(s) that is (are) the product of this cross, as well as host animals that are known heterozygotes and wild-type host animals. Probes to screen the Southern blots may be designed as set forth herein.

The knockout mammals described herein will have a variety of uses depending on the gene or genes that have been modulated. Where the targeted gene or genes modulated encode proteins believed to be involved in immunosuppression or inflammation, the knockout mammal may be used to screen for drugs useful for immunomodulation, *i.e.*, drugs that either enhance or inhibit these activities. Screening for useful drugs may involve administering the candidate drug over a range of dosages to the knockout mammal, and assaying at various time points for immunomodulatory effects of the drug on the immune disorder being evaluated. Such assays may include, for example, looking for increased or decreased T and B cell levels, increased or decreased immunoglobulin production, increased or decreased levels of chemical messengers such as cytokines (e.g., interleukins and the like), and/or increased or decreased levels of expression of particular genes involved in the immune response.

For example, patients undergoing chemotherapy often experience immunosuppression. It would be desirable to activate the immune system of such individuals by administering to the patient a therapeutic agent capable of producing such an effect. A knockout mammal as described herein could be used to screen a variety of compounds, either alone or in combination, to determine whether partial or total restoration or activation of the immune response results.

Similarly, the same strategy could be applied to find compounds that would be useful in suppressing the inflammatory response observed in many patients with arthritis, or useful in suppressing the autoimmune phenomenon observed in patients with rheumatoid arthritis and lupus. In addition, mammals may be useful for evaluating the development of the immune system, and for studying the effects of particular gene mutations.

In a preferred embodiment, the knockout mammals described herein are used for xenograft transplantation into human patients. The xenograft tissue may be from any mammal, preferably a pig. The xenotransplanted tissue may be in the form of an organ including, for example, a kidney, a heart, a lung, or a liver. Xenotransplant tissue may also be in the form of parts of organs, cell clusters, and glands including, for example, lenses, pancreatic islet cells, skin, corneal tissue, and the like.

In yet another aspect of the present invention, the target gene is the Gal  $\alpha(1,3)$  galactosyl transferase gene in pigs. The Gal  $\alpha(1,3)$  galactosyl transferase is an attractive target for knockout in the pig. This enzyme is responsible for the addition of a carbohydrate residue, Gal  $\alpha(1,3)$  Gal, that is recognized by human IgM and IgG antibodies in pig-to-human xenotransplanted tissue and leads to subsequent hyperacute rejection. Knockout pigs, which lack the Gal  $\alpha(1,3)$  galactosyl transferase gene, may thus potentially serve as a rich source for xenotransplanted organs. Nucleic acid sequences encoding Gal  $\alpha(1,3)$  galactosyl transferase and mutants thereof are disclosed. Preferably, the nucleotide sequence encodes pig Gal  $\alpha(1,3)$  galactosyl transferase. Nucleotide sequences may be in the form of DNA, RNA or mixtures thereof. Nucleotide sequences or isolated nucleic acids may be inserted into replicating DNA, RNA or DNA/RNA vectors (as are well known in the art), such as plasmids, viral vectors, and

the like (Sambrook et al, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, NY, Second Edition 1989).

Nucleotide sequences encoding Gal  $\alpha(1,3)$  galactosyl transferase may include promoters, enhancers and other regulatory sequences for expression, transcription and translation. Vectors  
5 encoding such sequences may include restriction enzyme sites for the insertion of additional genes and/or selection markers, as well as elements necessary for propagation and maintenance of vectors within cells.

Targeting constructs comprising nucleotide sequences, and mutants thereof, of the Gal  $\alpha(1,3)$ galactosyl transferase are particularly preferred as they may be used to inactivate wild type  
10 Gal  $\alpha(1,3)$  galactosyl transferase genes according to the methods of the present invention. Mutant Gal  $\alpha(1,3)$  galactosyl transferase nucleotide sequences include, but are not limited to, nucleotide deletions, insertions, substitutions and additions to wild type Gal  $\alpha(1,3)$  galactosyl transferase, such that the resultant mutant does not encode a functional galactosyl transferase. These nucleotide sequences may be utilized in the methods of modulating expression of  
15 galactosyl transferase of the present invention. In this manner, mutant sequences are recombined with wild type genomic sequences in the target cells.

In a most preferred embodiment, knockout pigs are produced in which the Gal  $\alpha(1,3)$ galactosyl transferase gene produces a non-functional protein. By producing a non-functional protein, the human antibody that would otherwise bind to the Gal  $\alpha(1,3)$ Gal epitope  
20 expressed on the xenotransplanted tissue does not bind, so that immune responses which give rise to tissue rejection are prevented. In this embodiment, any knockout construct capable of

modulating the interaction between antibodies directed to the Gal $\alpha$ (1,3) galactosyl transferase linkage may be used.

## EXAMPLES

The following examples are for purposes of illustration only, and are not intended to limit  
5 the scope of the disclosure or claims.

### EXAMPLE 1

Inactivation of the Gal  $\alpha$ (1,3) galactosyl transferase gene by insertion of an engineered active gene in the form of an engineered exon within intron 3.

In this example, the Gal  $\alpha$ (1,3) galactosyl transferase protein is prevented from being  
10 translated by insertion of an in-frame, promoterless engineered exon (*e.g.*, an antibiotic resistance gene) that contains multiple stop codons within an intron of the Gal  $\alpha$ (1,3) galactosyl transferase gene. Using this 'promoter-trap' strategy, the engineered exon is spliced in frame upstream of exon 4 of the Gal  $\alpha$ (1,3) galactosyl transferase gene. This results in the expression of the drug resistance gene prior to the gene of interest and concomitantly inhibits expression of the transferase gene due to the  
15 presence of multiple stop codons downstream of the drug resistance gene. As described herein, any gene that confers survival of the targeted cells under appropriate selection conditions may be used as the engineered exon, including, but not limited to, ampicillin, kanamycin, gentamicin, neomycin phosphotransferase, puromycin-N-acetyl-transferase, hygromycin b-phosphotransferase, thymidine kinase, and tryptophan synthetase. The present example employs neomycin.

20 A gene targeting construct is designed which contains a sequence with homology to the Gal  $\alpha$ (1,3) galactosyl transferase gene 5' intron 3 sequence, an intron 4 splice acceptor signal sequence, a promoterless neomycin phosphotransferase gene engineered to contain multiple stop



codons (engineered exon), the intron 4 splice donor sequence for splicing the engineered exon to the downstream exon 4, and additional intron 3 sequence homology to aid with annealing to the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene. Although this example describes targeting intron 3, it will be appreciated that the method may be used to target any intron within the Gal  $\alpha(1,3)$  galactosyl transferase gene or any other gene of interest. A sequence listing of the introns in the Gal  $\alpha(1,3)$  galactosyl transferase gene (from within intron 3 to the end of intron 8) is provided in Figure 1. A schematic diagram of the targeting vector and corresponding nucleotide sequence are shown in Figures 2 and 3.

The gene targeting construct is generated by ligating 5 distinct DNA fragments (1-5 below) together to form the final gene targeting construct using standard molecular biology techniques well known to those skilled in the art. The PCR reactions use the ELONGASE Enzyme Mix (Life Technologies, Gaithersburg, MD) according to the manufacturer's instructions. In the present example, a 50 ul final reaction volume is used, with 2 ul of DNA template, 1 ul of ELONGASE Enzyme Mix, 60 mM Tris-SO<sub>4</sub> (pH9.1) 18mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.2 mM MgSO<sub>4</sub>, 200mM dNTP mix, 10% DMSO and 200nM of each primer. The reaction is hot started at 95°C for 1 minute and followed by 30-40 cycles in a standard PCR thermocycler (GeneAmp PCR System 2400; PE Applied Biosystems, Foster City, CA).

1. A polymerase chain reaction (PCR) product consisting of intron 3 sequences as listed in Figures 1 and 3, nucleotide numbers 10-4020, is generated using standard PCR conditions for long range PCR of genomic fragments. Primers used include a 5' primer containing a NotI restriction site and intron 3 sequences 10-23 (GGCGGCCGCAGGCCTCACTGGCC); and a 3' primer containing a SalI restriction site and sequences homologous to intron 3 sequences 3999-4020 (GGTCGACGGATGCTGGGTGGAATAACAGG), where underlined sequence

indicates restriction sites and bold type indicates homology to endogenous sequences. An additional guanine nucleotide is added to the 5' end of all probes in this example to balance out 1 bp deletions that sometimes occur during cloning.

2. A PCR product is generated consisting of intron 4, the 3' splice sequence (the pyrimidine rich lariat and Gal  $\alpha(1,3)$  galactosyl transferase intron 3 dinucleotide acceptor sequences) and 196 bases 5' flanking the ag dinucleotide acceptor site (nucleotides 11521-11716 in Figures 1 and 3). Primers used include a 5' primer containing a SalI restriction site (GGTCGACCCACCGTTTGATCTGAG); and a 3' primer containing a EcoRI restriction site and the complementary strand homologous to the pyrimidine rich lariat and Gal  $\alpha(1,3)$  galactosyl transferase dinucleotide acceptor sequences (GGAATTCCTAAAAGCAAATGGAAATAAAAACATATC), where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence.

3. A PCR product consisting of a neomycin resistance gene (Genbank Accession #AF081957; Figure 4) is generated using a 5' primer containing an EcoRI restriction site, and homology to the neomycin resistance gene, including the ATG start codon (GGAATTCAATGGATCCCCACCATGGG); and a 3' primer containing a HindIII restriction site and complementary strand sequences to the 3' coding region of the neomycin gene, including the natural stop codon followed by two additional engineered stop codons (GAAGCTTCGGCTATTACTAAGTAGTGGATATCC), where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence (see Figures 3 and 4).

4. A PCR product is generated containing the 5' splice donor sequences for intron 4 of the Gal  $\alpha(1,3)$  galactosyl transferase gene, corresponding to sequences 4938-5173 in the claimed sequence comprising intron 4 (Figures 1 and 3). Primers used include a 5' primer containing a HindIII site and sequence identity to intron 4, sequences 4938-4962, including the Gal  $\alpha(1,3)$  galactosyl transferase dinucleotide splice site (GAAGCTT**GTAATTATGAAACATGATG**);
- 5 and a 3' primer containing a PstI site and complementary strand sequence from intron 4 corresponding to nucleotide numbers 5152-5173 and includes multiple stop codons (GCTGCAGCCACAGGTCACGGCAATGCGG); where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence.
- 10 5. A PCR product containing 1150 nucleotides of intron 3, corresponding to nucleotides 4024-4826 of the claimed sequence (Figures 1 and 3). Primers used include a 5' primer containing a PstI site and sequences 4024-4050 of the claimed sequence (GCTGCAGCCCTCTTCAACTACAATTT**CATGCAGC**); and a 3' primer containing a XhoI restriction site and complementary strand sequences to 4801-4826 of the claimed sequence
- 15 (GCTCGAGAGAAAATTAGATTAAATACACCCAGAG); where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence.

Each PCR fragment (steps 1-5) is separately amplified. A single PCR fragment is cloned into the pCR2.1 vector (Invitrogen, San Diego, CA) according to the manufacturer's ligation

20 instructions. The recombinant plasmid DNA is transformed into a suitable bacterial host (Invitrogen, San Diego, CA). The bacteria are cultured and plasmid DNA is isolated. Plasmid DNA with the correct insert, as determined by restriction analysis and sequence analysis, is used to construct the final product.

Following PCR fragment amplification, a series of ligations is performed to clone the final construct in the bacterial plasmid pBS SK+ (Stratagene, La Jolla, CA).

- a. The HindIII-PstI fragment from the intron 4 PCR product (step 4) and the PstI-XhoI 3' homology fragment from intron 3 (step 5 above) are ligated to a pBS SK+ vector DNA following digestion with HindIII and XhoI. The 3 DNA fragments are mixed in equal molar ratios and incubated in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid DNA is transformed into a suitable bacterial host (DH10B, Life Technologies, Gaithersburg, MD). The bacteria are cultured, and plasmid DNA is isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, is used to construct the final product.
- b. The resulting plasmid (step 5a) is digested with HindIII and EcoRI and ligated with the HindIII-EcoRI Neomycin resistance gene fragment (step 3), that has been previously digested with HindIII and EcoRI. The resulting recombinant plasmid DNA is transformed into a suitable bacterial host (DH10B, Life Technologies, Gaithersburg, MD). The bacteria are cultured, and plasmid DNA is isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, is used to construct the final product.
- c. The resulting plasmid (step 5b) is digested with EcoRI and NotI and ligated to the SalI-EcoRI intron 4-3' splice fragment (step 2) previously digested with SalI and EcoRI and the intron 3 4 kb NotI-SalI fragment (step 1) previously digested with NotI and SalI. The 3 DNA fragments are incubated in equal molar ratios in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid DNA is transformed into a suitable bacterial host (DH10B, Life

Technologies, Gaithersburg, MD). The bacteria are cultured, and the recombinant plasmid DNA is isolated.

5 This final construct is used to transfect porcine embryonic fibroblasts, transgenic pig fibroblasts, or porcine embryonic stem cells, or porcine primordial germ cells. Cell clones that are resistant to neomycin are screened by PCR to determine the site of integration. A primer located in the region of intron 4 not incorporated into the final construct (complementary strand of 5407-5427; GGACAATGGCAACATGGCAGG; see Figures 1 and 3) is used in combination with the 5' neomycin gene primer (step 3). Only targeted insertions yield the appropriate sized PCR fragment. All other integration events produce a negative result.

10 Cell clones with a targeted insertion are then used to generate transgenic animals using nuclear transfer techniques, or in the case of the stem cells, used to inject into developing blastocysts and produce chimeric offspring.

## EXAMPLE 2

15 Inactivation of the Gal  $\alpha(1,3)$  galactosyl transferase gene by replacement of exon 4 with an active gene in the form of an engineered exon.

In this example, the Gal  $\alpha(1,3)$  galactosyl transferase protein was prevented from being translated by replacing an endogenous exon (exon 3) with an in-frame, promoterless engineered exon (an antibiotic resistance gene) that contained a bovine growth hormone poly A sequence attached to the 3' end of the gene, which served to terminate transcription of the engineered exon.

20 The engineered exon was spliced in frame, so as to take advantage of the endogenous promoter typically used by the Gal  $\alpha(1,3)$  galactosyl transferase gene ('promoter-trap' strategy). This resulted in the expression of the drug resistance gene and concomitantly inhibited expression of the full-length

Gal  $\alpha(1,3)$  galactosyl transferase gene. As described herein, any gene that confers survival of the targeted cells under appropriate selection conditions may be used as the engineered exon, including, but not limited to, ampicillin, kanamycin, gentamicin, neomycin phosphotransferase, puromycin-N-acetyl-transferase, hygromycin b-phosphotransferase, thymidine kinase, and tryptophan synthetase.

5 The present example utilizes puromycin.

A gene targeting construct was designed which contained a sequence with homology to the Gal  $\alpha(1,3)$  galactosyl transferase gene 3' intron 3 sequence, an intron 3 splice signal sequence (splice acceptor sequence), a Kozak consensus sequence, a promoterless puromycin N-acetyl transferase gene linked to a bovine growth hormone poly A sequence (bpoly A) (engineered exon),  
10 the 5' intron 4 splice signal sequence (splice donor sequence), and a sequence with 5' intron 4 sequence homology. Exon 4 of the Gal  $\alpha(1,3)$  galactosyl transferase gene codes for ATG start codon and the N-terminal portion of the protein. Although this example describes targeting introns 3 and 4, it will be appreciated that the method may be used to target any exon within the Gal  $\alpha(1,3)$  galactosyl transferase gene or any other gene of interest. A sequence listing of the introns in the Gal  $\alpha(1,3)$   
15 galactosyl transferase gene (from within intron 3 to the end of intron 8) is provided in Figure 1.

The gene targeting construct was generated by ligating two distinct DNA fragments together to form the final gene targeting construct using standard molecular biology techniques well known to those skilled in the art. The first DNA fragment was obtained from the 3' end of intron 3 containing the 3' splice sequence (the pyrimidine-rich branch site used in forming the  
20 lariot during splicing and the AG dinucleotide splice acceptor sequence). The second DNA fragment was obtained from the 5' end of intron 4 containing the GT dinucleotide splice donor sequence. The fragments were ligated into the pBluescript vector containing a Kozak consensus sequence in-frame with the coding sequence of a promoterless puromycin gene linked to the

bovine growth hormone poly A sequence (Figure 5) to form the final gene construct. A schematic diagram of the targeting vector and corresponding nucleotide sequence are shown in Figures 6 and 7. Additionally, this construct has been deposited with ATCC on 28 September 2000 with accession number \_\_\_\_\_.

5     1.     Generation of the first DNA fragment.

The first DNA fragment was a polymerase chain reaction (PCR) product consisting of intron 3 sequence as shown in Figure 7 (nucleotide numbers 235-4851, positions relative to nucleotide position 1 of the insert isolated from the lambda phage clone) and generated using standard PCR conditions as described by Randolph *et al.*, (1996) for long range PCR of genomic  
10   fragments. The 5' primer, consisting of intron 3 sequences 235-260, was

5'-AAGATTATAAATAGCCTCGTGTCAGG-3'. The 3' reverse primer sequence was complementary to sequence 4827-4851 at the extreme 3' end of intron 3 and containing the AG splice acceptor site, and was 5'-CTCCTGGGAAAAGAAAAGGAGAAGG-3'.

PCR reaction conditions to generate the 4.616 kb intron 3 sequence were performed using  
15   the ELONGASE Enzyme Mix (Life Technologies, Gaithersburg, MD) according to manufacturer's conditions. In the present example, a 50 µl final reaction volume was used, with 2 ul of DNA template, 1ul of the ELONGASE Enzyme Mix, 60 mM Tris SO<sub>4</sub> (pH 9.1), 18 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.2 mM MgSO<sub>4</sub>, 200 mM dNTP mix, 10% DMSO and 200 nM of each primer. The reaction was hot started at 95°C for 1 min, followed by 30-40 cycles in a standard PCR machine  
20   (e.g., Gene Amp PCR Systems 2400; PE Applied Biosystems, Foster City, CA).

2.     Preparation of the PCR2.1 cloning vector.

The Not I site of the PCR2.1 cloning vector (Invitrogen, San Diego, CA) was destroyed to avoid carrying over a second Not I site into the final construct. The Not I site was unique and

used to linearize the final plasmid construct. The PCR2.1 vector was digested with Not I and the overhangs filled-in using the Klenow enzyme (Roche Molecular Biochemicals, Indianapolis, IN) according to the manufacturer's specifications. The plasmid was re-ligated using T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations.

- 5 Plasmid DNA was transformed into a suitable bacterial host (Top 10 F', Invitrogen, San Diego, CA). The bacteria were cultured, and plasmid DNA was isolated and incubated with Not I enzyme to confirm loss of this site by restriction analysis.

3. Insertion of the first DNA fragment into the PCR2.1 vector.

- Following PCR, the 4.616 kb fragment was ligated into the modified PCR2.1 vector  
10 using T4 DNA ligase according to the manufacturer's specifications. Plasmid DNA was transformed into a suitable bacterial host (*e.g.*, Top 10 F', Invitrogen, San Diego, CA). The bacteria are cultured and plasmid DNA is isolated. Plasmid with the correct insert in the proper orientation, as determined by restriction analysis and sequence analysis, was used to construct the final product.

15 4. Preparation of the second DNA fragment.

- A 2.084 kbp PCR product consisting of the intron 4 homology sequence containing the GT dinucleotide donor consensus splice sequence was constructed using standard PCR conditions as described above in step 1. The 5' primer consisting of sequence 4938-4961 at the extreme 5' end of intron 4 was 5'-GTAATTATGAAACATGATGAAATG-3'. The 3' primer  
20 was homologous to the complementary strand of intron 4 at position 6997-7021 and has the sequence 5'-AGCCAGCGCTTACTAAGTACGTTGC-3'

5. Insertion of the second DNA fragment into the PCR2.1 vector.



Following PCR, the 2.084 kb fragment was ligated into the pCR2.1 vector from Invitrogen (San Diego, CA) using the manufacturer's ligation conditions. Following ligation, the recombinant plasmid DNA was transformed into a suitable bacterial host (Top 10 F', Invitrogen, San Diego, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, and orientation in the plasmid, as determined by restriction analysis and sequence analysis, was used to construct the final product.

6. Preparation of a synthetic oligonucleotide linker sequence.

A synthetic oligonucleotide linker containing a Kozak consensus sequence and relevant restriction enzyme sites was prepared for in-frame cloning of the promoterless puromycin gene:

10 Xho I | Kozak seq.---| HpaI | Hind III| Bgl II | Sal | Eco RV | Eco RI  
TCGAGCCACCATGGTTAACAAGCTTAGATCTGTCGACGATATCG  
CGGTGGTACCAATTGTTCTGAATCTAGACAGCTGCTATAGCTTAA

7. Assembly of the gene targeting construct.

The following ligations were performed to generate the final construct in the bacterial plasmid pBS KS+ (Stratagene, La Jolla, CA). The final construct is illustrated in Figure 6:

a. The oligonucleotide linker containing the Kozak consensus sequence (step 6) was ligated to the pBS KS+ vector DNA following digestion with Xho I and Eco RI. Ligation was carried out using at least a 3:1 molar ratio of linker to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Stratagene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Restriction enzyme analysis was performed to confirm successful ligation using unique

restriction sites within the linker (Bgl II or Hpa I). This plasmid containing the linker was then used to construct the final product.

b. The resulting "mother" plasmid (step 7a) was then digested with Eco RV and Spe I to clone in the 3' arm of the targeting construct. The 2.084 kb PCR fragment cloned into the PCR  
5 2.1 vector (step 5) was digested with Eco RV and Spe I, isolated away from vector DNA by agarose gel electrophoresis and purified. The 2.084 kb fragment was ligated between the EcoRV and Spe I sites of the mother plasmid (step 7a). Ligation was carried out using a 3:1 molar ratio of insert to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid  
10 was transformed into a suitable bacterial host (XL1-Blue MRF', Strategene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was then used to construct the final product.

c. The next fragment cloned into the mother plasmid (step 7b) was the cassette containing  
15 the promoterless puromycin gene coding sequence with the bovine growth hormone gene polyA signal sequence attached to its 3' end following the TGA stop codon (Figure 5). The PGK puromycin bpolyA plasmid (used as a positive control for puromycin resistance of transformed cells) was digested with Hind III and Xho I. The puromycin bpolyA fragment was separated away from the rest of the vector DNA containing the PGK promoter by electrophoresis on a  
20 0.7% agarose gel and purified. The mother plasmid (step 7b) was digested with Hind III and Sal I. The Hind III/Xho I puromycin bpolyA cassette was ligated to the Hind III and Sal I sites of the mother plasmid. Ligation was carried out using a 3:1 molar ratio of insert to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the

manufacturer's recommendations. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Strategene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was used to construct the final product.

5 d. The final cloning step involved ligating the 5' arm of the construct, which was the 4.616 kb intron 3 insert from the PCR2.1 vector (step 3). The PCR2.1 vector (step 3) was digested with Kpn I and Xho I. The 4.616 kb PCR fragment was isolated away from vector DNA by agarose gel electrophoresis and purified. The 4.616 kb Kpn I/Xho I insert was ligated into the mother plasmid (step 7c) that was digested with Kpn I and Xho I. Ligation was carried out using  
10 equimolar ratio of insert to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Strategene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated by standard molecular biology techniques. Plasmid with the correct insert, as determined by  
15 restriction analysis and sequence analysis, was used as the final product.

e. The final construct may be used to transfect porcine embryonic fibroblasts, transgenic porcine fibroblasts, or porcine embryonic stem cells, or porcine primordial germ cells. Cell clones that are resistant to puromycin may be screened by PCR to determine the site of integration by methods well known to those of skill in the art. A primer located in a region of  
20 intron 4, which is not incorporated into the final construct, may be used in combination with a 5' puromycin gene primer. Only targeted insertions will yield the appropriate size PCR fragment. All other integration events will produce a negative result.

Cell clones with a targeted insertion may then be used to generate transgenic animals using nuclear transfer techniques, or in the case of stem cells, used to inject into developing blastocysts and produce chimeric offspring.

### Example 3

- 5 Inactivation of the Gal  $\alpha(1,3)$  galactosyl transferase gene by replacement of exon 4 with an a reverse orientation active gene.

In this example, the Gal  $\alpha(1,3)$  galactosyl transferase gene was functionally inactivated by using a "collision construct" to insert an active gene in place of an exon and at least part of the flanking introns, including the splice donor and splice acceptor sites. The inserted gene is under the control of a highly active promoter such as the phosphoglycerate kinase I (PGK) gene promoter, such that transcription of this gene causes the termination of transcription of the endogenous gene (Rosario *et al.*, (1996) Nat. Biotech.14:1592-1596). Exon 4 of the Gal  $\alpha(1,3)$  galactosyl transferase gene codes for ATG start codon and the N-terminal portion of the protein. Thus, the insertion was made to replace exon 4 as well as a portion of the flanking introns 3 and 4, resulting in a truncated transcript that did not code for a functional enzyme. Although this example describes targeting introns 3 and 4, this method could be used to target any introns within the Gal  $\alpha(1,3)$  galactosyl transferase gene or any other gene of interest. A sequence listing of the introns in the Gal  $\alpha(1,3)$  galactosyl transferase gene (from within intron 3 to the end of intron 8) is provided in Figure 1.

In this example, the PGK promoter was inserted driving the expression of the puromycin resistance gene with the bovine growth hormone poly A (bpolyA) transcription termination sequence. This gene replaced the Gal  $\alpha(1,3)$  galactosyl transferase exon 4 as well as a portion of the flanking intron 3 and 4 sequences by standard homologous recombination techniques utilizing intron 3

and 4 sequences for homology flanking the inserted gene. Intron 3, which separates exons 3 and 4, is greater than 5 kb in length, and a construct was built such that there was at least about 4.6 kb of homologous sequence on one end of the gene. Intron 4, which separates exons 4 and 5 is about 6.8 kb in length, and the construct was built such there is at least about 2.2 kb of homologous sequence on the other end of the gene. Positive selection for transfected cells in which the construct has been integrated was accomplished via expression of the puromycin resistance gene. As described herein, it will be appreciated that any selection marker gene that confers survival of the targeted cells under appropriate selection conditions may be driven by the strong PGK promoter. Additionally, a toxin gene was inserted to eliminate random integration events.

10 The collision construct was generated using standard molecular biology techniques well known to those skilled in the art. The 4.616 kb intron 3 homology fragment and the 2.084 kb intron 4 homology fragment were generated using PCR and cloned into the PCR2.1 cloning vector as described in Example 2, steps 1-5 above for the replacement targeting construct. The generation of the collision construct first involved ligating the 2.084 kb intron 4 homology  
15 fragment into the pBS KS+ vector as the 3' arm of the collision construct, followed by the PGK-puromycin-bovine polyA cassette in the opposite orientation to the coding sequence of the GT gene. The 4.616 kb intron 3 homology fragment, as the 5' arm, was cloned in next. This generated the targeting construct for homologous recombination. The ricin A toxin gene was also added to the plasmid outside the region of homology, which will effectively kill a  
20 percentage of the cells in which random integration has occurred. The ricin A toxin gene was PCR amplified and cloned based upon the published sequence (Figure 8). A schematic diagram of the final construct is shown in Figure 9. Additionally, this collision construct has been deposited with ATCC on 28 September 2000 with accession number \_\_\_\_\_.

1. The 2.084 kb 3' arm of the construct (intron 4 homology fragment) was the first fragment to be ligated into the pBluscript cloning vector, which was modified to contain the XhoI – EcoRI linker (see Example 2, step 6) within its multiple cloning site (pBS KS+). The ligation of the linker into the vector is described above in Example 2, step 7a, and ligation of the 2.084 kb 3' arm into the Eco RV and Spe I sites of the vector is described above in Example 2, step 7b.
2. The next step involved ligation of the PGK puromycin bovine polyA cassette into the pBS KS+ vector, which contained the 2.084 kb 3' arm. The PGK-puro-bPA cassette was digested with Eco RI which was immediately 5' of the PGK promoter. The Eco RI overhangs were blunted by filling in with Klenow enzyme (Roche Molecular Biochemicals, Indianapolis, IN) using the manufacturer's specifications. The PGK-puro-bPA cassette was then released from the vector by digestion with Xho I, which was immediately 3' of the bovine polyA sequence. The blunted PGK-puro-bPA-Xho I cassette was separated from vector DNA by agarose gel electrophoresis and purified. The pBS KS+ vector (step 1) was digested with Hpa I and Xho I, and the blunted PGK-puro-bPA-Xho I fragment was ligated between the Hpa I and Xho I sites of the vector. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Stratagene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was then used as the final product.
3. The 4.616 kb intron 3 homology fragment was ligated into the pBS KS+ mother plasmid (step 2) and represented the 5' arm of the collision construct. Isolation of this fragment from the PCR2.1 cloning vector and ligation into the Kpn I and Xho I sites of the mother plasmid is described above in Example 2, step 7d.

4. The ricin A toxin gene (Figure 8) was inserted into a commercially available mammalian expression vector, *e.g.*, pcDNA1/Amp (Invitrogen). The insert was then excised with the CMV promoter and the SV40 poly A site and cloned into the Not I site of the recombinant plasmid by blunt end ligation, following Klenow fill in reactions on both the insert and vector.
5. Following ligation, the recombinant plasmid DNA was used to transform a suitable bacterial host (XLI-blue, Strategene). The bacteria were cultured, and plasmid DNA isolated. This final construct DNA was then linearized with Kpn I (a unique enzyme site in the plasmid MCS outside of the construct sequence). Linearized plasmid may be used to transfect porcine embryonic fibroblasts, transgenic porcine fibroblasts, or porcine embryonic stem cells, or porcine primordial germ cells. Cell clones resistant to puromycin may be screened by PCR to determine the site of integration by methods well known to those of skill in the art. A primer located in the region of intron 4 not incorporated into the final construct may be used in combination with a 5' puromycin gene primer. Only targeted insertions yield the appropriate size PCR fragment. All other integration events produce a negative result.
- 15 Cell clones with a targeted insertion may then be used to generate transgenic animals using nuclear transfer techniques, or in the case of stem cells, used to inject into developing blastocysts and produce chimeric offspring.

#### EXAMPLE 4

- Isolation of porcine genomic DNA encoding the Gal  $\alpha(1,3)$  galactosyl transferase gene from a  
20 Lambda phage clone library.

In this example, a pig genomic library was screened using a cDNA corresponding to exon 4 of the Gal  $\alpha(1,3)$  galactosyl transferase gene using molecular biology techniques that are well

known to those skilled in the art (*e.g.*, Sambrook *et al.*, *supra*). A pig genomic library (Clontech, Palo Alto, CA) was obtained and screened with a PCR fragment derived from exon 4 of the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene. Exon 4 was labeled with  $^{32}\text{P}$  dCTP using the Random Prime Kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. Approximately 4 million phage forming units were screened and unique clones that contain exon 4 sequences as determined by Southern blotting were isolated. Clones obtained by this procedure contained inserts 15-40kb in length. These clones, designated pgGT, lambda 1, lambda 2, lambda 4-1 and lambda 8-2. Five vectors comprising unique, overlapping nucleotide sequences which span the entire the pig Gal  $\alpha(1,3)$  galactosyl transferase gene from within intron 3 through intron 8 have been deposited with the ATCC: (1) a 1.6 kb insert within intron 3 of the extreme 5' end of the 18.275 kb lambda-2 phage clone, (2) a 6.7 kb HindIII fragment spanning intron 3 to intron 4 of the 18.275 kb lambda-2 phage clone, (3) a 4 kb HindIII fragment following the 6.7 kb fragment 2 of the 18.275 kb lambda-2 phage clone, (4) a 6 kb HindIII-SalI fragment at the 3' most portion of the 18.275 kb lambda-2 phage clone, and (5) a 13 kb fragment of the lambda-2 phage clone spanning exon 7 to exon 9. These five vectors were deposited with ATCC on 29 September 2000 with accession numbers \_\_\_\_\_, respectively. Subclones of the various inserts were used to generate the claimed intron sequences from within intron 3 to intron 8 as provided in Figure 1 using molecular biology techniques well-known to those skilled in the art (see *e.g.*, Sambrook *et al.*, *supra*). These sequences may be used to determine regions of sequence homology in design of targeting constructs for modulation of the pig Gal  $\alpha(1,3)$  galactosyl transferase gene

Although the compositions and methods provided herein have been set forth in detail, one skilled in the art will recognize that numerous changes and modifications may be made, and that



such changes and modifications may be made without departing from the spirit and scope thereof.

We Claim:

1. A method of modulating the expression of a eukaryotic gene in a cell, comprising  
  
transfecting said cell with a nucleic acid construct, said construct comprising a first  
construct sequence homologous to a first gene sequence, a sequence encoding a selectable  
5 marker, and a second construct sequence homologous to a second gene sequence, wherein said  
first and second gene sequences independently comprise at least a portion of one or more intron  
regions of said eukaryotic gene, and  
  
integrating said selectable marker into said eukaryotic gene,  
  
wherein expression of said selectable marker results in modulation of expression of said  
10 eukaryotic gene in said cell.
2. The method of claim 1, wherein said first construct sequence and said second construct  
sequence are each homologous to at least a portion of an intron region of the gene.
3. The method of claim 1, wherein said sequence encoding a selectable marker is integrated  
into said eukaryotic gene by homologous recombination, wherein said first construct sequence  
15 recombines with said first gene sequence and said second construct sequence recombines with  
said second gene sequence to insert the selectable marker into the gene.
4. The method of claim 1, further comprising screening said cell for expression of said  
selectable marker.
5. The method of claim 2, wherein said first construct sequence and said second construct  
20 sequence are homologous to different regions from within the same intron of the gene.

6. The method of claim 2, wherein said first construct sequence and said second construct sequence are homologous to regions of different introns.
7. The method of claim 1, wherein said selectable marker gene is an antibiotic resistance gene.
- 5 8. The method of claim 1, wherein said sequence encoding a selectable marker is a nucleotide sequence which, when expressed, confers a phenotype selected from the group consisting of ampicillin resistance, kanamycin resistance, gentecin resistance, neomycin resistance, puromycin resistance, hygromycin b resistance, thymidine kinase activity, tryptophan synthetase activity, adenine phosphoribosyltransferase activity, dihydrofolate reductase activity,  
10 and histidinol dehydrogenase, anthocyanin, beta-glucuronidase and luciferase.
9. The method of claim 8, wherein said sequence encoding a selectable marker confers neomycin resistance or puromycin resistance.
10. The method of claim 1, wherein said eukaryotic gene is selected from the group consisting of genes encoding B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2, VCAM-1, CD28,  
15 CD80, CD86, CD154, major histocompatibility complex class I B-2-microglobulin, invariant chain, caspase-1 caspase-3, and Gal  $\alpha(1,3)$  galactosyl transferase.
11. The method of claim 10, wherein said eukaryotic gene encodes Gal  $\alpha(1,3)$  galactosyl transferase.
12. The method of claim 11, wherein said Gal  $\alpha(1,3)$  galactosyl transferase gene is a porcine  
20 gene.

13. The method of claim 12, wherein said first construct sequence and said second construct sequence are independently selected from homologous regions of the intron selected from the group consisting of intron 3, intron 4, intron 5, intron 6, intron 7, intron 8, and intron 9 of the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene.
- 5 14. The method of claim 13, wherein intron 4 has the nucleotide sequence of nucleotides 4938-11716 in Figure 1.
15. The method of claim 13, wherein intron 5 has the nucleotide sequence of nucleotides 11753-13748 in Figure 1.
16. The method of claim 13, wherein intron 6 has the nucleotide sequence of nucleotides  
10 13810-14358 in Figure 1.
17. The method of claim 13, wherein intron 7 has the nucleotide sequence of nucleotides 14463-21627 in Figure 1.
18. The method of claim 13, wherein intron 8 has the nucleotide sequence of nucleotides 21766-27048 in Figure 1.
- 15 19. The method of claim 13, wherein said first construct sequence and said second construct sequence are homologous to different regions within the same intron of the eukaryotic gene.
20. The method of claim 19, wherein said intron is intron 3 of the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene.
21. The method of claim 13, wherein said first construct sequence and said second construct  
20 sequence are homologous to different introns of porcine Gal  $\alpha(1,3)$  galactosyl transferase gene.

22. The method of claim 21, wherein said first construct sequence is upstream of said second construct sequence.
23. The method of claim 21, wherein said first construct intron region is homologous to an intron 3 region and said second construct intron region is homologous to an intron 4 region of  
5 porcine Gal  $\alpha(1,3)$  galactosyl transferase.
24. The method of claim 2, wherein said sequence encoding a selectable marker is a promoterless gene.
25. The method of claim 2, wherein said sequence encoding a selectable marker further comprises a promoter.
- 10 26. The method of claim 25, wherein said promoter is a phosphoglycerate kinase (PGK) promoter.
27. The method of claim 2, wherein said sequence encoding a selectable marker is transcribed in the opposite orientation relative to the orientation of said eukaryotic gene.
28. The method of claim 27, wherein said sequence encoding a selectable marker further  
15 comprises a promoter sequence.
29. The method of claim 1, wherein said cell is selected from the group consisting of a fibroblast, epithelial cell, endothelial cell, transgenic embryonic fibroblast, embryonic stem cell, and primordial germ cell.

30. The method of claim 2, wherein said cell is a porcine cell.
31. The method of claim 2, wherein said construct further comprises an AG dinucleotide splice acceptor site.
32. The method of claim 2, wherein said construct further comprises a GT dinucleotide splice donor site.
33. A nucleic acid construct comprising a first construct sequence homologous to a first gene sequence, a sequence encoding a selectable marker, and a second construct sequence homologous to a second gene sequence, wherein said first and second gene sequences independently comprise at least a portion of one or more intron regions of a eukaryotic gene.
34. The nucleic acid construct of claim 33, further comprising an AG dinucleotide splice acceptor site.
35. The nucleic acid construct of claim 33, further comprising a GT dinucleotide splice donor site.
36. The nucleic acid construct of claim 33, further comprising a Kozak consensus sequence.
37. The nucleic acid construct of claim 33, wherein said sequence encoding a selectable marker is a nucleotide sequence, which when expressed, confers a phenotype selected from the group consisting of ampicillin resistance, kanamycin resistance, gentecin resistance, neomycin resistance, puromycin resistance, hygromycin b resistance, thymidine kinase activity, tryptophan synthetase activity, adenine phosphoribosyltransferase activity, dihydrofolate reductase activity, and histidinol dehydrogenase, anthocyanin, beta-glucuronidase and luciferase.

38. The nucleic acid construct of claim 37, wherein said sequence encoding a selectable marker confers puromycin resistance or neomycin resistance.

39. The nucleic acid construct of claim 33, wherein said eukaryotic gene is selected from the genes encoding B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2, VCAM-1, CD28, CD80, CD86,  
5 CD154, major histocompatibility complex class I  $\beta$ -2-microglobulin, invariant chain, caspase-1, caspase-3, and Gal  $\alpha$ (1,3) galactosyl transferase.

40. The nucleic acid construct of claim 39, wherein said eukaryotic gene is porcine Gal  $\alpha$ (1,3) galactosyl transferase.

41. The nucleic acid construct of claim 40, wherein said first construct sequence and said  
10 second construct sequence are independently selected from homologous regions of the intron selected from the group consisting of intron 3, intron 4, intron 5, intron 6, intron 7, intron 8, and intron 9 of the porcine Gal  $\alpha$ (1,3) galactosyl transferase gene.

42. The nucleic acid construct of claim 41, wherein intron 4 has the nucleotide sequence of nucleotides 4938-11716 in Figure 1, intron 5 has the nucleotide sequence of nucleotides 11753-  
15 13748 in Figure 1, intron 6 has the nucleotide sequence of nucleotides 13810-14358 in Figure 1, intron 7 has the nucleotide sequence of nucleotides 14463-21627 in Figure 1, and intron 8 has the nucleotide sequence of nucleotides 21766-27048 in Figure 1.

43. The nucleic acid construct of claim 41, wherein said first construct sequence and said second construct sequence are homologous to different regions within the same intron of the  
20 eukaryotic gene.

44. The nucleic acid construct of claim 43, wherein said intron is intron 3 of the porcine Gal  $\alpha(1,3)$  galactosyl transferase gene.
45. The nucleic acid construct of claim 41, wherein said first construct sequence and said  
5 second construct sequence are homologous to different introns of porcine Gal  $\alpha(1,3)$  galactosyl transferase gene.
46. The nucleic acid construct of claim 45, wherein said first construct sequence is homologous to an intron 3 region and said second construct sequence is homologous to an intron 4 region of porcine Gal  $\alpha(1,3)$  galactosyl transferase.
47. A cell transfected with the nucleic acid construct of claim 33.
- 10 48. A cell transfected with the nucleic acid construct of claim 41.
49. A cell transfected with the nucleic acid construct of claim 44.
50. A cell transfected with the nucleic acid construct of claim 46.
51. A bacterial cell transformed with the nucleic acid construct of claim 33.
52. A bacterial cell transformed with the nucleic acid construct of claim 41.
- 15 53. A bacterial cell transformed with the nucleic acid construct of claim 44.
55. A bacterial cell transformed with the nucleic acid construct of claim 46.
56. A nucleotide sequence of intron 4 of the Gal  $\alpha(1,3)$  galactosyl transferase gene having nucleotides 4938-11716 in Figure 1.



57. A nucleotide sequence of intron 5 of the Gal  $\alpha(1,3)$  galactosyl transferase gene having nucleotides 11753-13748 in Figure 1.
58. A nucleotide sequence of intron 6 of the Gal  $\alpha(1,3)$  galactosyl transferase gene having nucleotides 13810-14358 in Figure 1.
- 5 59. A nucleotide sequence of intron 7 of the Gal  $\alpha(1,3)$  galactosyl transferase gene having nucleotides 14463-21627 in Figure 1.
60. A nucleotide sequence of intron 8 of the Gal  $\alpha(1,3)$  galactosyl transferase gene having nucleotides 21766-27048 in Figure 1.
61. A lambda phage clone derived from a porcine genomic library comprising at least a portion of  
10 the Gal  $\alpha(1,3)$  galactosyl transferase gene, wherein the lambda phage clone is selected from the group consisting of pgGT, lambda 1, lambda 2, lambda 4-1 and lambda 8-2.
62. A method of making a transgenic mammal comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 33, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating said embryo into an  
15 appropriate host mammal, and generating offspring from said impregnated host mammal.
63. A method of making a transgenic mammal comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 44, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating said embryo into an appropriate host mammal, and generating offspring from said impregnated host mammal.

64. A method of making a transgenic mammal comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 46, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, and generating offspring from said impregnated host mammal.

5 65. A transgenic mammal made according to the method of claim 62.

66. A transgenic mammal made according to the method of claim 63.

67. A transgenic mammal made according to the method of claim 64.

68. A method of reducing transplant rejection comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 32, selecting for transfected cells comprising the nucleic acid of the  
10 construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.

69. A method of reducing transplant rejection comprising transfecting a nuclear donor cell with  
15 the nucleic acid construct of claim 44, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.

20 70. A method of reducing transplant rejection comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 46, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate

host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.

71. The nucleic acid construct of claim 43, further comprising a nucleic acid sequence  
5 encoding a gene which is toxic to said eukaryotic cell

72. The nucleic acid construct of claim 71, wherein said gene which is toxic to said eukaryotic cell is the ricin A toxin gene.

1/27

**Figure 1**

Sequence from within Intron 3 to the end of intron 8

```

1  gtcgactcta  ggcctcactg  gcctaatacg  actcactata  gggagctcga  ggatcaatta
61  gaggtccacc  atccctttcc  tgaatgccta  aggccagata  tgttggaatt  tagaattttt
121  caaatgcaga  atatttatcc  tatattatgt  aacgccccca  gtgcagcaac  agcacataac
181  aatgtacatc  aatatttatg  caaagaaatg  tttaaacagt  ctactaagt  gataaagatt
241  ataaatagcc  tcgtgtcagg  gcctggatgc  caactgaatt  ataaacaggc  ttttggtttt
301  cagagcttgg  agttggatga  gggctctgaga  aactgctcca  tgttcagggt  taccagttct
361  gtgggtgtct  ccagacccca  cctccttccc  aagctctctc  accaccacac  cttctctggg
421  agtgaagaca  acggcagaga  ggcatggcca  cagtggccac  agtctccacc  ccgatctgtc
481  tgctcccaaa  cccaggcctt  tcctcgcact  cagtgcctaa  gctgttgatg  taggagtcaa
541  gtggcttttt  ccagcatctg  ggccgagctg  catgtagccc  cgtgcatttc  gtaactttgc
601  cctgggcccc  gggctgtttg  tgccaggacc  tgagggtgctc  acaggaaaag  agtcccatct
661  ccccatcttc  tcaccatctc  tggaaacacca  cctatcatga  ttgtatctga  aagggtggcg
721  ttgaatcaga  ggtttccaaa  cagagctcac  tttccaagca  agaaggaata  gagtgcacatg
781  gctgataatc  ccatactttc  tcttctttta  ctggatttca  caacagaggt  gatggagcac
841  ctgagatcta  agcctggagt  caccctcaga  ccctctctgc  aaatatattg  agaataaccc
901  gtcccctgaa  aggacacatc  tcagtgccat  tctcatttca  ttcacacatc  tttttttttt
961  tttttttttt  ttttttttgg  gcttttttgc  atttcttggg  ccagtccctg  ggcataatgga
1021  ggttcccagg  ctaagggtct  aattggagcc  gtagctgcag  gcctacgcca  gagccagagc
1081  cacacgggat  ctgagccgag  tctgcaacct  acaccacagc  tcacggcaac  gccggatcct
1141  taagccactg  agcaaggcca  gggatggaac  ccacaacctc  atgtttccta  gtcagattcg
1201  ttaaccacag  agccacaacg  ggaactccca  cacattattt  attgcggcc  ttctctgtct
1261  tctgtggggc  actgggaatt  caggggtgat  caagaagtca  tccctcctgc  cctcaggaag
1321  ctcaaaccac  tcattattta  ttgacggcct  tctctgtctc  ctgtggggca  ctgggaattc
1381  aggggtgacg  aagaagtcac  ccctcctgcc  ctgaggaagc  tcaaacaagc  aggtagagga
1441  ggcagagcaa  aatgcaggtc  ttatccggtg  agccgactcc  cagggcgatg  tgtacagcaa
1501  aggaatagag  ggatgggggc  cggaggagag  aaaagggtct  cagccgtggt  caggggtggg
1561  gtgggaagtg  gcttcacaaa  ggcagtgaca  ttggctccca  ggtgtccact  cttctgtctc
1621  tgctaccttc  tggctccttc  cttctggggc  ctctctctat  ctacctctaa  agcttcaccc
1681  acatcctcct  ttctttttct  ctctctggat  tctctcctgg  gtaatcaaat  tegtccctt
1741  cacgtcagat  ccgatacgtt  ccttgggtcca  tgaacaactt  ctccgattgc  atgggtctgc
1801  tacatctctc  tgatgaactt  tagacttgaa  tgtccacttg  tctccctgtc  cctttttagg
1861  tattcgaca  ctccccgaca  ttcacacgtc  caaaaggga  ttcatgatta  ttatcctcca
1921  agcctgttcc  tcctccagcc  catctgagaa  aatactacaa  cccccctgt  taatcagaaa
1981  tcttgggtct  tccttgtctc  atctctgata  acaaaattac  caaccacgtc  ctatcaattc
2041  tctctccaaa  gtatatatat  atatatattt  ttttaatttt  ttcccgctgt  acagcatggg
2101  gatcaagtta  ttcttacatg  tatattttcc  ccccaccctt  tgttccgttg  caatatgagt
2161  atctagacat  agttctcaat  gctactcagc  aggatctcct  tgtaaatata  agttgtatct
2221  gataacccca  agtcccgat  ccctccact  ccctccctct  cctgtcgggc  agccacaagt
2281  ctattctcca  agtccatgat  tttcttttct  gtggagatgg  tcatttgtgc  tggatattag
2341  attccagtta  taagtatat  catatgggat  ttgtcaaagt  atatatatta  tttttctttg
2401  tctttttgtc  ttttgtcttt  tttttgttgt  tgttgttgtt  gttgttgttg  ttgttgctat
2461  tacttggggc  gctcccgagg  catatggagg  ttcccgagct  aggagttaa  tcggagctgt
2521  acccacgggc  ctacgccaga  gccacagcaa  cgcgggatcc  gagccgcgtc  tgcaacctac
2581  agccacgctc  acggcaacgc  tggatcctta  acccactgag  caagggcagg  gaccgaaccc
2641  gcaacctcat  ggttcctagt  cggattcgtt  aaccactgag  ccacgacggg  aactcccaaa
2701  gtatattttg  aatcaagcca  ccctttgagc  caggccacct  cctctttatg  gtcagtgaag
2761  cgggtctgcc  ttgtcctttt  ctccattctc  cacactcagc  acccagatgg  gtctctctag
2821  gtgaagttgg  atcaggggat  tctccagctt  tagatgcttt  ttgggattcc  ccaccctact
2881  ttccatacct  ttccaggttc  tgactgcctc  tgccccctt  ctgactgcct  agcaccagcc
2941  actcaagggg  gacagtgtca  gtcactattt  tttctctgtc  caggtttttt  gcttttgttt
3001  ttttcaaaaa  cgagcagctc  tttctcttgt  ctgcctggta  tagatgctgt  ttccaaaata
3061  ttctcatccc  ttctcacggc  ccttgtcatc  ctttcccatc  ctatcttcat  cccttgggaa

```

2/27

```

3121 gctctaaagt catctcccca aattgaaggg tgactaaaga gtttcccaga agggaaaaact
3181 gagtttccaa ctactacact gacttgcaag aaatgtttgt gtcttcatta aatgaaaaag
3241 aaaaaactgt aacaagatat gagaaaatac agaaaggaaa taataagact agaaaagtca
3301 aatataatagt gaaggtgttg catcaaacac ttaaataaac tagtacagat gttaaaagac
3361 taaattatat agttgaagga tagctgtgaa gatgtaaact atgacatcta aaacacaaaa
3421 tgttggcggt cccgtcacgg cacagtggaa acgaatccga ctaggaacca tgaggttgca
3481 ggttcaattc ctgcccttgc tcagtgggtt aaggatccgg tgttgccgtg agctgtgggtg
3541 taggtagcca atgaggcttg gatcccgctg tgctgtggct ctggtgtagg ccggtggcta
3601 cagctccgat tcgacccta gcctgggaac ctccatagc cggggagcg ggcccttaaa
3661 aagacaaaag accaaaaaaa aaaaaaaca aaaaaccac aaatgttgg gaatcagtc
3721 tctactagta ttatgttatt gtcaagtttt ctttttatgt ctgttaatat ttgcgttcta
3781 gatgtagggt ctctgatata gtgtgcatat atgttaacca atgttatgtc ttctctgggt
3841 attgatccct ttgttattat gtaatgccct actttatctt ttgttacatt ctttgtttat
3901 gagtattgct gatattgtgg tagctgccac acttttcttg tctttccat ttacaataaa
3961 tatctttcta tctccacca aattaaagta ctccgcaacc tgttattcca ccagcatcc
4021 cttccctctt caactacaat ttcattgcag gatcaagaaa tagaatgtac cgactgtttg
4081 ccattggtgg ggcattggaa aagtgggtgg aaagtgcaga gcttagatta taaaggccag
4141 ggtgagagtt cccattgtgg tcagctgaaa tgaatctgac tagcatccat gagcacgaag
4201 gtttgatccc tggcctcaat cagtgggtta aggatctggc gttgctgtcc gtgagttgtg
4261 gtgtagttcg cagacaaggc gtggacttag tgtggctgtg gctgtggcat aggttagtgg
4321 ctacagctct gattcgacct ctgacctggg aatctctata tgctgtgagt gtggccctaa
4381 aatttaaagt aaattaaata aaggaccagg gtatatttt ctttggaggat aaggtacata
4441 gtcagtatat caaggacagt agacctagga aacggatgct tctctagtc tgtgatgcca
4501 ggtggggcat ctgagttggg ggcggttggg gcccttaggg accattaact aaaccgctca
4561 ctctcccaca tctcggtgga ccttgggatc agtcaggatg cttcccttt gagcctcaaa
4621 atggccttag tatecttccc aaccagacg gccctgtcag ttcattgact tggctaattt
4681 gccagtgtag gctatgcaa attaaggtag aacgcactcc ttagcgctcg ttgactatc
4741 atcaactttt ctttttagaa aagatattgg tataagcact tcttaaaaaa ccatattcca
4801 ctctgggtgt atttaatcta attttccctt ctctttttct tttcccagga g-----

```

```

-----EXON 4-----
4938 -----gta attatgaaac atgatgaaat gatgtgatg aaagtctcct
4981 ctaatctcct agttatcagc caagtcacca gcttgcatga aaagtaggat tcactgacac
5041 cgtaaagaaa gcattccaga gagttgccgt tgtggctcag gggcagcaaa cccaattagg
5101 atccaagagg aggtgggttt gatccctggc cttgctcttt ggcttaagga tccggcattg
5161 ccgtgacctg tgggttaggt tgcagatgca gctcggatct ggcattgctg tggctgtggc
5221 gtaggctggg ggcttcagct ccagtttgac ccctagcctg ggaacttcca tatcccacac
5281 ttggggccct aaaaatcaaa gaaagaaaga aaatattcta cccttctgt atccctgagc
5341 ccttaaatat cgtctttaaa gtcattagat cttcaagtac cttccagcta attaatatc
5401 ttccttctcg ccatgttgcc attgtcctga tttttatacc tctgcagttc tgggtaggct
5461 agagccagaa ataataaggt catgttaaga ccaagatata atattaaatt atttatatga
5521 ccagatatgg aagttacctt gagaactttc agacaggaat tccatgagaa atacaccctg
5581 atttttgcaa tctaaaaata tttgcagagt ttaaagggaac aactcaagtt gttgactttt
5641 gctgcaaaac aactgagtc gctggtgatt catattgtgc tggctaaact ttgggtgtt
5701 ttgtcttttt ttttttttaa ctctggaag caaatgaat taaacatttc tgagttttca
5761 aattcatcag tggattcacc ccaaatattt gagctgcttc tttgcttttg gaaactacga
5821 tgccttggag attccagctg gagacgcttc tgacagaaag aaatgtctgc aagcagctac
5881 aaaaatgcat gatggctttg acttaagagg cattgatacc gcttggcctt tctttcaaaa
5941 aggccacctt acaacttggc ctgaaggcat tcccgtcgtg gtgcagcgga aaatgaatct
6001 gactaggaac cccgaggttg tgggttcaat ccctggcctt gctcagtggt ttaaggatcg
6061 ggtgttgaa taagctgtgg tgtagattgc agacgcagct tggatctggt gttgctgtgg
6121 ctttggtgta ggccggcagc tacagctcca cttggacccc tagtctggga accttttagg
6181 tgtggcccta aaaggaaaaa agacaacaaa caaacaaaaa accaaaaaac aacttggcct
6241 ggagagctat gtcattacca ttgatatttt gatgggtagt gtttttagtag cccctcaagt
6301 tcaggatgat ggcttgatt aacattagaa tgtctcttaa attctacgac ttgatgagcc
6361 agcaggacca ttttggccac ttagaaagga actgcactct caggtccatc agtagaagga
6421 ggattctcta gggagtctc tcttagctca gcgggttcaa gaattcagtc ttgtccctac

```

3/27

6481 agcagctcag gtgactgcta tggtttggtt ttgatccctg gcccaggaat ttctgcatgc  
 6541 tgcaggtgca gccaaaaaaaa aaaaaaaaaa aaggaggagg nggattccct agaataagaa  
 6601 gctatcattc ctttggatgc ttcatagatc taaccacttc tggaacagtt attccctctc  
 6661 attctgaaga actcatttta agaaaaacaa gacgagctag agagtgaaca aatgggtctac  
 6721 aaaccaggcc ttctgaattg aggaaactgt ggtacttcct ctgaagaaaa gatgacagcg  
 6781 ttggatgcag agaccctggg gctcccttag gtacttgagg actgaggaga tattctcagt  
 6841 ggaggctgga gctaggctgc ctggggctgg tcctgtgcca ccacttccct cctctgtgac  
 6901 tttgggcaag ttccctatc tttaaaaatg gggatgatag tagtacctgc ttcataagggt  
 6961 tggtggataa aataagttgt gaataaagca ctaagggcaa cgtacttagt aagcgctggc  
 7021 tgccatcacc accaccacta tcaccatctg tccggagggc agcataggac aggaattttt  
 7081 tggcaaatag aaggaaagat tctaggagtt cccgttggtg tgcaggggaa atgaatccaa  
 7141 ctaggaaacta ggaggtttcg gggtcaatcc cgcgcctcgc tcagtggtt aaggatccag  
 7201 tggtgccatg agctgtggtg tagattgcag acatggctag gatctggagt tgctatggct  
 7261 gtggtgtaag ctggcagctg tagctcggat tctacccta gcctgggaat ttccgtatgc  
 7321 cacaggtttg gccctacaaa gaaaaaagaa aaagaaaaag aaaaaattct aggggctgaa  
 7381 agaactaac agaagagcaa gttcccatg gggttcctga cctgagttga gatgcttggtg  
 7441 taggcaacct tcaagctctg aactcttgat tgttttgaat tgcagccaga gttgtacttc  
 7501 catattttgg gtacttcaca aaattaaaaac acagaagcca aaggcccaga agtgcatatt  
 7561 ggtgctggcc tcccataaag aggggtggtt tgcagtgtg ggcacactct ctcttcacag  
 7621 taactggagc agattctggc tgctcttcag ggccgtagtc tggcaccag actgcagcca  
 7681 catcattctt caatgtgagg aatctatttg aacatctgca aggggtttaa aaggcaggag  
 7741 attctttgcc acctgtgtaa ttggtctgag gtgagctgag ggcactaac tttagacggt  
 7801 gggtagcact gtagctaaag aggattacag gagtccctgt tgtggcttag tggtaacaaa  
 7861 tccaactagt atccatgagg attcagggtc gatccctggc ctgcgtcagt gggtcaggta  
 7921 tccggtgttg ctgtggctgt ggtgtaggct ggcagcttca ttttatttac ccctagcctg  
 7981 ggaacttcca tgtgctgtag gtaaggccct tgaaaaaaa aaaaaagaga tttcaaaata  
 8041 actccatcaa acacatacag ctgtttaaga atgtcatcca ggacagcatt tggttaaagg  
 8101 ctagatgaaa aaaaaaaaaa aaaaacttag aattttattt atttattttt tcttttttagg  
 8161 gccagacctg tggcctatgg aaatgcctgg gctaggggtg gaatcagagc tgcttacacc  
 8221 acagccatag ccacgccaga tccaagcccc gtctgtgacc tacaccacag ctcatggcaa  
 8281 aactggatc cttaatccac tgagtgaggc caggaattga acccacattc tcatggatgc  
 8341 tagttgggtt cttaagccac tgagccacaa gcttagaatt ttagagggtg aagaaacttt  
 8401 aagagctata ataaagtaat gatggtgtag gtgatttga tgttagcggc taactagttat  
 8461 attgtgttg cttgtgccag gaactccact gttcattccc tctgtttttt aaacagccc  
 8521 tggaaagtca gtgttagtcc acatttctag atgaggaata ctgagtttcc acaatattaa  
 8581 atgtgaacgt tcaaggtcac atttttagga agatttaggt ccagggtgt ctgacttggg  
 8641 taacctgggt aaccttccct ttagtcaagg gttccattgt tcaggcgatg aaatggagac  
 8701 ccagtaggtg aaatgactta acagtgaact tatgtccaac ttctaattag aactcagatc  
 8761 ttctgattca tcatctgggg ctcttctctg agctggttgt tgatgcbaaa tgctgcgagg  
 8821 ggtacagtgt gccgtcaagg agaatcccta ccctcaaggg gttatgctgt agatggagca  
 8881 ggcagaggta cccatgaaag cccaacaaca caggctagaa ggaggatgtc agagagagag  
 8941 agcaaggaa cgtgagagtt caggggaggc aagattatgt ttggcttggga gatggatcta  
 9001 tgttttgcat ttattttttt gggggggggg tctttttgct acttcttggg ctgctcccga  
 9061 ggcataatga ggttcccagg ctagggtctc aattggagcc gcagccacca gcctacacca  
 9121 gagccacagc aacgcaggat ctgagccgag tctgcaacct tcaccacagc tcacggcaac  
 9181 nccgggatcgt taaccactg agcaagggca gggaccgaac ctgcaacctc atggttcccta  
 9241 gtcagattcg ttaagcactg cgccacgacg ggaactccct catttagaaa tatttattga  
 9301 gcacctactg tatgccaggc attgtgctag gttcatacca aagaaggctc agaagagatg  
 9361 gcatccgagc tgtgccttga aggatgaata tgtgttaaat gccgtacact tcagggtggt  
 9421 tggtgctgtg acctgaggtg ttgaaggctt ctgggaaagg aggggtgagat gaggaagagg  
 9481 gaggggttac taaaaagatg ggacgaggtg gcaaatccaa atctataaat tgatgccctg  
 9541 agtgccctgc aggagggtgg ggctcctgag tgctgggtgg caggggccct tccccctcct  
 9601 cttgccccct tcccttcccc ctctttagag atctgaagtc agattcccca ggttcaaata  
 9661 ctggttcttc ccttagcagt atgaccttgg gcaaaataat ttattgcctc tgcctctctg  
 9721 aggaggaata gaacctcctt cattgactgt tattagaatt taatgagcta atacatgtca  
 9781 gttgcttaga aagggtccca gccaaactatt agctattatg aatattatca gatcaataga  
 9841 cagatttaga aacaagggac tttagagctg ggtccatggg tactgagctt agaggggaaa

4/27

9901 ccataggtgg taggaaggca tgtatttcat tcctaccagg agatgtggac tcccagctgg  
9961 ggcagaaggc agagggagga gatcggggct ttggcagaat ctcaaacaaa tattagtggg  
10021 tagtggtttt ttgtttctgt tttaaagat gagggcaggc gtttccgatg tggcgagtg  
10081 gaaacgaatc tgactagtat ccatgaggat gcagggtcca tccctggact cactcagtga  
10141 gttaaggatc cggcattgcc gtgagttgtg gtgtnggtca cagacacagc tcagatctgg  
10201 cattgctgtg gctgtgggtg aggctagcaag ctgtagctcc aattcaacc ctagcctggg  
10261 aacttccaca tgccgcaggt gcaaccccaa aagataaatg aataaataaa taaatatgcg  
10321 accttccttt cttggggcct ttgcatgttt ttctctctgt taggcacact cttgctaata  
10381 cctcttcaact gggcctccta agtatccttc agaactcagc taaaacatca tcccctcccc  
10441 tggggagcct tcgaggtcct cctgttaagt gtcctatgc tttcttggag ttttgaagtc  
10501 ctataatgat gtgtttatca aaatagggtc caccctccct gccagcttct ctacaccaca  
10561 gacacatggg gtctgtttca gtcaacactg tatgtctggc acttgacatg taacgcatgc  
10621 tcagcaggta tttgttgaat gaatggaggc ggtctgctag agtcgtcata tatttactga  
10681 tcccgctctg taggatgggc tcactgcttt tgtagctta agaagtacct tttttttttt  
10741 tttttttaat ggccacaccc atggcatata gaaattccac gaaggaagga agaaagaaag  
10801 aaagaaagaa ggaaattcct gggtcaggga ttgaatccaa gccacagggt caacctgagc  
10861 tgcagttgtg gcaacaccac atcctttaac ccactgtgct gggccaggga tcatacctgt  
10921 gcatgtacag cgacccaagc cagggcagtc agattctttt tcttcttttc tttctttctt  
10981 ttcttttctt tttttttttt tttttttttt ggctttttgc cttttctagg tgcggcatat  
11041 ggaggttccc aggctagggt tcgaatcaga gctgtagacg ccggcctaaa ccacagccac  
11101 agcaacacag gatccaagcc ttgtctgtga cctacaccac agctcacggc aacgctggat  
11161 ccttaaccgg ctgagcgagg ccagggattg aaccgcaac ctcatgggtc ttagttggat  
11221 tcgttaacca ctgagccatg atgggaactc ctgcagtcag attcttaacc cactatgcca  
11281 cagcaggaac tcctagaagt gccctttgag gctactctgt agacagctct gagccagcga  
11341 ggcaagacct gtttttctgg aggaagataa atcctgggtg agggatgggt gggctgtggg  
11401 cttcctggga cccatctctg gagcctctct ccctcagcaa agccacctg gacaataaga  
11461 gctgccatct attttttttt ctttaaaacta agatttgata ttttccagag acctcccctc  
11521 ccaccgtttg atctgagtaa ttctgaaatg acgagagtc cgtgatata ttttttcgat  
11581 ctgcaagggt gaaacctggg agtagccaca acccaggctc tcagctcagc ctagggtttc  
11641 aatgataatg attgcaaat agcttttctc tgcattccaa gtaacatgat atgtttttat  
11701 ttccatttgc ttttag---- ----- EXON 5----- ----- --gtaagtgc  
11761 tttgaattcc aaatatctct aggtcacctt ccatgtgacc ctggtggccc tacagtccat  
11821 tcttaacatg gcaggtgggt acgcacttgt ggtcctaggt ggaggagagg gatgggggtc  
11881 caggggtctg agctgtactt ctccagcccc tagacttgcc tttctagagc atgagttgtg  
11941 tttttccttt gcttctcatc aagtatctat ctctttaagt gatgttgttt ggagaacatt  
12001 cctgccttgc tcataaaaaa gaatcagagt agatattatc cattatgcta cctactacat  
12061 gtgtataaaa gacccttgcc agaaattttg ccaagacaaa ggattaggaa gaaaggctgg  
12121 gtgtcctgat aaactagtgt gtgtattatt attatttatt attattacta ttactgggtga  
12181 ttaaggactc taagccttca tttttctttt tttttttttc ctatcttcga cttgggtgct  
12241 aggaagctta gagcaaagta ttgtgcttaa atgcttgcct tttccttggc cttcattttt  
12301 tttaaaacat tttttcttat taaagtatag ctgatttata gtacacctca tctgatatga  
12361 tttatcccct ggtgttaaat cctggccttt gttagatgcc atgggatctt ggcaatttgc  
12421 tcaaaactcat tttgccaata tcttagctat gaagtaaaaa taaagttaaa gattttgttc  
12481 tcacagagtg gctgggatga ccaaagtcac gtgaaaacac ccgagtgact aaaatgtttc  
12541 tctgtttcgt tttgttttgt tttgattctt gtattgtttt cctattttatc gtaaccacac  
12601 tttcttcata agccatttca agcacttctt gaaagtagat ggactttaag tttcttggac  
12661 ttccagttgt ggtgcagtgc aaacaaatct gactagtatc catgaggatg catcttcgat  
12721 ccctggcctt gctcagtggg ttaaggatct ggtgctgctg tgacctgtgg tgtaggtcac  
12781 agaggcggct cagattccaa gttgctgtgg ctgtggcata ggcgggcagc tacagctcca  
12841 attagacccc tagcctggga acctccatat gccacgggtg cggccctaaa aaaacaaaaa  
12901 aacaaaaaaa aataataaaa taaaaaaaag taagtttctt actactatgc ctattcctgc  
12961 ctaccacact catttgtcca accaaaagggc atggcagctc ctggctttcc cctcttgtc  
13021 cctcctctag acggtccacg ccttggagag actctcacag tgagtgtgga catctgagca  
13081 tgtgtgtgga gaaaaacggc ttctgattac agtttgcgtg gcttcgggtt tagaaacctt  
13141 caggttttct gaatctgtca ctgctgacct ttgtagcaac tccagttctc ctcatcttagc  
13201 aagagatcat tcttggaat gtgggctttt cttttttaag actacaactt actacagtaa  
13261 ttttcagctt ataaagctct tttacgtagg ctactttgtt tagtcttcga agccatccat

5/27

13321 cttcatggta ggtttggtgc tgtcattgtt ctctaacttt ccagaggagg caatggtgat  
13381 tcagagagga aaaatgtcct atcagggtta cacaactgga aaggggcaga tccaggactt  
13441 ggatttatgg attagtgcct tcccccttgt cattcttctt ccccccttct ctcctgcca  
13501 gggctccctg gaccagctga gtgccggggg ccacccaagg gaggggagat ctggggagtg  
13561 tcgggtccat cctctgtcat tcagctggag cagcttcatt tctgttcatt ttacgaatta  
13621 gtgatttttt aaaaaatgat ttaagtgtat ctgccataaa aagagtttg aacaccccc  
13681 accccctgcc ccagtttggg actgcctcat tttctaaaaa atgtttgtca tcttttcat  
13741 ttcttag--- exon 6---  
13810 ----- gtaagactgg gaaacggcca tctgtgtatc tgctcaaggc tgtagagtcc  
13861 aaataaaatg gtttcacagc catgaccttc tcagtcgcct cgtccttctg gcatattgga  
13921 atcatgcaag gactctgttg cgtgggtggg atgggcatta agtttggttt tcaagggttt  
13981 gttttttccc atttgatgtg ggagtgtgtg aggtcaggc tgaagttgag gctgcaggaa  
14041 aggtgtcat gagcttacat acgtgataa ggactgtttt atcagggttc cctgtcctg  
14101 acagaacctg agaattgctg ggtcattaca ttgctaactg tgctggctctg gaaggatgat  
14161 gaaaaagggg agcactgaac cttttcatca gacggacgtc ttacgaata aatgtttagt  
14221 atcacaagcc ttctgtagac atttgcttgg cgcacgtgtt ctctgccttg gaatctgttt  
14281 tcacaagatc ctcactctca ttctggcaca tgggtcaccc tccctgcctt cctcagcttg  
14341 ttttccgttt gtacgtag--- EXON 7---  
14463 ---gtaagaa agaagcggt gccctatttc agtaaatcca cagctcagca ggaggtggca  
14521 ggggctgtgt aggaacacag gccatgagaa cattcccgac aggggagaca gagttttagc  
14581 atctattaga cccttaaaac agggctcggg tttgtgtgt gtgtgtgtgt gtgtgtgtgt  
14641 gactggcgtg gcatcccggt aggtggggag tctggatgtt ttgactcaaa tgtttgcgca  
14701 acaacatagt gattctgaca gcaattagcc aatggcaaca gatgaatcac aaaatactgg  
14761 ggctagcagt tcccggtgtg gcacagcggg aatgagtcca agtaggaacc gtgaggtttc  
14821 acgttggatc cctggcctca ctcagtgggc taaggatccc gcgttgccat gagcgggtgt  
14881 gtaggttgca gacgtggctc ggatctgtgt ttgtgtgtgt aggccggcag ctgtagctcc  
14941 agtttgacct ctacgtggg aacctccata tgctgcgggt gcagccctaa aaagcaaaaa  
15001 aatactgggg ctaggaaagg tcctatagcc catggggtgc tgtggaaagc cacagaatct  
15061 agaatacagat aaagaaaatt cttaggttcg gttaactcct ctgtaaaatg gcaatgacgc  
15121 gagttcccat tgtggttcag tggaaacaaa cccaactagt atccattcta tccccagccc  
15181 cactcagtgg gtttaaggatc tggcatttgt gcaagctgtg gtgtaggtca aagacaaggc  
15241 tcagatctgg tgttgcgtg gctgtgtgt agactggcaa ttacagctcc gattcaaccc  
15301 gtacactggg aacttccata gctgtcacct gtggccctaa aaaaataaaa ataaaataaa  
15361 atggcagtga tgatttaaaa aaccttaca ggccatgcac acactctaag caaacataca  
15421 acaatgttca ttaaacaccc tatgatatgc tagcccatgt gtgtcagact atcatttgat  
15481 ccattaatgt gtttagaaac tggcttaagg gtcaaaaatt gaaaccaa ataggaaatcag  
15541 aaggcatctc atgaagtaag gataagtatt gtttagtggc atttttgctt ccatatgtgt  
15601 gtgtgcatgt gtttgtgacg taggatatat ttactatga gtcacggtac tagacattgt  
15661 gtttaggcaca ggagacagac aggagaaatt aatcaataca cttatattca agaagtttac  
15721 agtcctatag gggcactaga tgtgaaacta gcatttgcca agattggtaa ggcgggtgtga  
15781 ttaacctgac ttggaagaat gtggaaggga gtgaggaggt gggggtgtc cgggatgact  
15841 gggaggcctg agagactggt gcatgctggt gccgttgacc aggaaaggga ctggaggagg  
15901 agatacaggt ctaggaatag gggagggtgat gagcttgatt aaggatgtag gtgattcag  
15961 gtacctgtgg ggcatttggg tagagttcat ggagagccta tcttagaccg tcagactact  
16021 gtaacagaat accatagacg ggggtggtta taaacaacaa aaatgtattt tgtacagttc  
16081 tggaggctgg gaagtccaag atcaaggcac cagcagatgt ggtgtctggt gagaaccac  
16141 tttctagttt atacatgccc gtcttctcat tgtgttctta ggccatggaa ggggagaggg  
16201 agctctccag ggcttctttt ctaagggcac ccatccatt cctgagggtt ccaccctcag  
16261 gatctaata cccaatgcc ccacccttg aggttgggtt tcaacacacg aatttggggg  
16321 gggacataat cgtcagtgtg cagcagggcc ttaagctgg cagaagatat gtggtagaga  
16381 cattggaaga ggagccatca gagttcaggc cgaggagtag agccaagtct tatctgacat  
16441 cagcacttaa agggctcagg gagggaggaa aagaagtctg aaacggagtg ggagcagcag  
16501 tagaggtagg agggagatcc gaggagaata gcatctattc cagcatgagg ggtcttgga  
16561 tctggaagga ctgagtatgt tctggcaagt ttcttactgt tttaggtagg ggaggtccat  
16621 gtcaagtgtg aaagaatggg aatttggccc tcccatgtg ttatttgagg aggtggcac  
16681 agggtagcca tctatggaat gctaagctac ttctgggaag tggagaatg tggaaaacgc  
16741 tatacctttt ttttttcttt tgagggggca cacctgcagc atatggaagt tcccaggcta



6/27

16801 ggggtctaataat cggagcttca gcagcttgcc tacaccacag ccacagcaac gcaggatccc  
16861 caacccactg agtgaggcca gggatcgaac tggcaacctc atggatacta gttggattca  
16921 tgtctgctgt gccatgacgg gaactcccac cagagctttg ataagcccct tagatgtgtg  
16981 tgcattgtggg gaggtgtgag actccccagg ggccagatag ttcagctaga gacagtcacc  
17041 aagctccctt caagagagag tggcacactg actgtggtac ttgtgcttta ctgagtggga  
17101 cttgaaaggt ggagccaggg ctgaatcact ggggtcagag gtgaaagacc acaagccacc  
17161 cactgtcact cttttccact ttcccccatg catggctcag tttctcctcc aaatgaatgt  
17221 tgagcagata tttaaagaat attgatattg agtagcacta aatctaagcc ttctcagggc  
17281 accaaaagta aattgcttgc ataaatgcag gaaagatgag ccactgaatt aatataattta  
17341 tagaaatata ttatagatat atttatagaa atgtgtcctc tgtattctac ctgaaacata  
17401 ttaagtgtat tctgtataca tttttacatc aagatctctc agcctctgta gcacccacat  
17461 cctcttagtt ttctccgac ttctccagct ccttgtcag tctccttgat tggggcccct  
17521 cctccttctc agccagggaag tgttagagtt tctggagggt gactccatta tctcatctc  
17581 ttttctttat acactgtctt catccatttc tgtggcatta gataccattc atgggacaat  
17641 gactactgca tttatatctt tagcccaaag gttttctctg agcccccagt tgtgtctaac  
17701 tcttgccctga catcactgca ggtctcgcat gaccagatg gaggttgggt ttttttgtt  
17761 gtttttttgt tttttgtctt tttaggccgt acctgaggca tagggagttt cccaggctag  
17821 ggggtgaatt ggagctctag ctgcccggcct gcaactgtac cacagcaaca ccagatccaa  
17881 gccgtgtctg tgacctacac cacagctcac ggtaacgcca gatccttaac ccactgagca  
17941 agggcagggg tcgaacctgc atcctcatgg atgtagtca gatttgctcc cactgagcca  
18001 tgacgggaac tcccagaggg agtttttgat tccacccac cctcattctt ctcatttccc  
18061 atctcaagca agagtcagga gacctccatg aattcactgt tcatccttcc ttcattccagt  
18121 ccacagtcta attcatctac atgttctgtt cattcttccc acgtcatctc cctctctctc  
18181 atcacccctt tccatggcat cattagctct ttgtggatc actgtgtacg tctatagct  
18241 gtcacctgga atggcaatcc caggcctaga gtctaacaca ttgttgcat cctcattttt  
18301 ttccacctgc tgtatgattc atagaatttt aacactttcc tctgcccagc ccttcaggac  
18361 ccacaagtgt actcataaat aagtacataa aattgaaaag ctaccaacta cctaggtagg  
18421 tgggtgccgtt tggatttttg agaccctttt attttgtgga agtcttttgt ttgcaactgg  
18481 ggaacccccc atgccgaagt aacgcagtggt gtgcccggct taaaaaactg agcagtttcc  
18541 agtgttttcc ggacttttcc ctctttttcg agccgtctct tgatgacgtt gggcaagctc  
18601 ttgctcttac acatacttgg cctgcttctg gttacttctt cttttctctg gggccggggg  
18661 ggggggtggg tttgagcaga ggttcatcca ggaagctgat ctcaatcatc ctcccacctc  
18721 ctctctctgc ccacggccgg tgttttcagc ctgtcatatg gcttccctag cagggttgt  
18781 gacatgttgg aattgccagt ggaatttaat aattaacagc tgggaaatag agcattttta  
18841 tttttaattc aacgttaatt ttttaatttg ctacttttcc tccccctgc ttaaaaggaa  
18901 tcttcccagc caagtaacca aactggacac tgggtgtctg aatcggtctt tctagaaggt  
18961 gtagtctgag ggggtgtgag gctgaggagt gtgcccagc ctgaccacct gtcagggatg  
19021 ggaagtctc aaggggtagg atgagtctga tctccgactt tagccctggg ccccgccaga  
19081 tggagcaggg acttggcaga tgtgcatcga ggagccgatg ccttggtgtg gcccctcta  
19141 ttggttgaat ggctaaaaac atagtttcct tctctctcac ccccaacttt ctttatttgt  
19201 gaaatggagg tgatgatagc acccactctt aaatgtggtt gtaaggattc aatgagagaa  
19261 taaaagtccg gtacttaacc ccagtgttg cgactagtaa ataagaatac ctcttctct  
19321 cctcttctc cttctcttct tcttcttctt ctttgcttaa agataatatt gctgcttgc  
19381 cagatgttca gctcacacac agctgcattg cttaactgat aggtgagggt ctgaactgct  
19441 aagagatcct tgacaggcta cactaaggac aggaatgaat gaagtgcagc tgcattggaa  
19501 acggttatag aatgataagc ctaagctgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtt  
19561 ttggtttgaa atccagtcac tcaacaaagc ttgggctttt atgtgagaaa agcatcacat  
19621 aagtttctaa aaacaaaaca aaacaaaaca aaaaaagct catcattntt ttttattctg  
19681 ccctgataaa aatgctgagg tagctcaaat gagtggaaat gatgacatac ctgggtttta  
19741 actccagttc attggtacag caagtaataa gccaaagaag catagagggn agtggttcag  
19801 agcacagacc acagtcagac tatctgggtt caaatccan cttggtctct tactantgt  
19861 gtgacctcag acaagttatt tcaccactct gngctacagt ttctcagag ataaanggga  
19921 atcaaaatat agcaacccca cccccacaaa aggtggaagc tcttggtctt gtgctgtgtc  
19981 ccataggaag gagccagtgt gtgttagtta ccatcatctg ctgtttggcc ttggtcaagt  
20041 taacaagcct tcgtgagcct tgattttcat atctgtacat tgagcatac acatggccaa  
20101 gcttttgcaa ggactaaatg agagaaagt tatgtgaaag tacctggcac aagttggtac  
20161 agagtcacaa acagcttctt tcatgattgt tggaaaccaa gagcagggtg ccccaagggg

7/27

20221 cgggtcatacc tgggggtgtga aggcagagct ggggtgtggg gtttgagcag cctagccatg  
 20281 tcccctccat ctctgcagaa ttactcttag gggaggaagg aagcaggctt gaagcagggc  
 20341 atgggggtggg ggtgggaggg cagaggccct ccagaatacg ggtgcagagc ttggtgacga  
 20401 gggacagaag gcatgagcaa gaagtggggg tgaccagcc ttgcccacc tcgcaggga  
 20461 acatggtttc tcccatgggg gaggggaggg tttgtgaaaa cccttcctcc acccgctat  
 20521 gaggggtgggg gtatgtttct gctgggctaa catctcctag acagcctcac cagggccata  
 20581 cccaggcagc cagcaggggc ccgatgggac cccaagacg ttgccccaac cactggcttc  
 20641 tacctgcagt ggctgaaata caagcacaca gctagcaggg gccaggccct ctcatgttgt  
 20701 gctggggggtt gaatctaaat gaccacgttt cagtctgttt ccgacgggga agtgcagcca  
 20761 cttccagagc aaatgggggt tgagccaaagt cttctcagct ccctcctgcc tctgtggggc  
 20821 cccagctgct acccagactt cccgcattta ctctttacaa tttgtccgtt gctcttccgt  
 20881 ggcttcctgt gtttgtgttg gcgttgctaa gtccctttca agtgttcttg gtgcagagaa  
 20941 catgccccct cccctccgcg ttttgtgctt ctagtcraga ggcgacatga tctgagagag  
 21001 gtctgagccc gaccctgtg tgtaacccat ggccacagtt agcccgccgg gtcccttctt  
 21061 ttctgggtga ctgtctcctt cttcgtaaaa tgaaccacaa caccaggtcc ccttagtctg  
 21121 gggccctttt tgtccgataa ccaggttcct acccagaaat tgctttcttg cagaaggcaa  
 21181 actgagacag cttcttcctc tttcagctca aatgtcactc tctcatcccg ctagtcaagc  
 21241 catagggtt tctcagggtc aggtggcacc cactgggata agtaacaccc aaagatgtcg  
 21301 ctggcagctt aggaaggggc tggggagata ggcaaggggg tttggaagga agattgagga  
 21361 cgaggacagc agaacagggg gacggaaggt acatgcatgt tgtacaggtg cgatcccaa  
 21421 agggggccac agggcagccc tcagaggcac ctggggcaga gcctcctgtc cctcccagc  
 21481 aagatgctgc aatgtcacac caccagctga ctggggctaa aatacagtca ggtatccagg  
 21541 ccagtcacca caagccatga ctgaccatg tcccccaga ctgtcgtacc ttagcaaagc  
 21601 catcctgact ctatgttttg tcaccag--- **EXON 8**-----  
 -----  
 21781 aaaactggcc ttgagttttt ccccttcac tatcagagga tgggtgaggg gccctgggt  
 21841 ttacagagc tgttcatgtc atgtctgaat tagtgagag gagaatggtg tcacagggc  
 21901 atttttagact cccttctgct gaggtcccca aaggctaaga ataaaactag tcagagggtc  
 21961 aactctttcc cacctcaggt gaggggcttg ggtgagggg aagaagatct gctataccca  
 22021 ctgcaccag agtcgatcga cagtacacc acagccacct ccgacctgac ctccacggc  
 22081 ctctgtggaa attccaaaaa tggcaattgt tagatggcct gtgtgtgct cgtttctct  
 22141 atgctctcga gacccagac caaggaccaa agacagaagt gtcctaagt gagggtttt  
 22201 ccagctccgc atggctgacc actcctctgt gccttctgcc tattcctctt gagggtttg  
 22261 gcccggggag attaggtgct tcagaacttc ctcttcaga ggttggatag gttcctgtt  
 22321 cagccccctg ttcttgtttt ggaactttt tccccaaact gtaaacattc ttacttaaaa  
 22381 gtagtagctg taatgttctg ttaaaatata acccagtttt cttttttaga gaatccccc  
 22441 cccttttttt ccaaaacaaa agcaaaagtt caattttct gttcacctcc gtgccccctc  
 22501 cctccccac atccatgggc ctttccatct gtacctttt tgaaagccac aaagaaactg  
 22561 aatcactttg ttatacagaa aaaattatca caacattgta agacaactat acttcaacaa  
 22621 aacttaaaaa aaaaatcaga aagagaaaga aagaaaaaga aaaaaaggaa ggaaggaagg  
 22681 gagtcccat catagctcat tgggttaatga atctgactag catccatgag gacacaggtt  
 22741 caatccctgg ccttgatcag tgggttaagg atctggtgtt gccgtgggct gtggtgtagg  
 22801 tcacagatgc atttcgaatc ccgcatgtg gtggtctga cgcaggccat catcacagct  
 22861 tcgattggac ccctagcctg ggaggtgca gccctaaaaa agcaggaaag gagaaaggga  
 22921 aggaagaaag aaaaggaagg aagggaaga gaaagagaaa gagagaaaga gaaaggaaga  
 22981 gaaaggaaga aaggaagaaa gaaaggaaga aagaaagaaa gaaaggaaga aagaaagaaa  
 23041 gaaggaagga aggaaggaag gaaggaagga aggaaggaag gaaggaagga aagaaagaaa  
 23101 gaaagaaaga aagaaagaaa gaaagaaaga aagaaagaaa gaaagaaaga aagaaagaaa  
 23161 gaaagaaaga aagaaagaaa aagaaagaaa aagaagttag ttgacctgtt gcctgtacaa  
 23221 agagaagtga aggtcaaggt tgatctgagt gaagatttaa gcgtcctct gagaccttct  
 23281 ccaacctaga gcccagaagg tatcactgtc actgtactaa tcaagtccta gtgtccctga  
 23341 atgtactgat cccaggggccc tgggtggcat ctctaataga gaaggggtgac tctggagttt  
 23401 tgacctttcg actagaagaa tatgttctgt taggagttgg gaaagattcc cagctcacct  
 23461 aatcctttgg tataagagga gaccagcag ccatactgca gattgattta acctgggtg  
 23521 ttccaaaaaag aacggctcca cattggctct agccatcttg ttgcagttt agagaagtaa  
 23581 actgtacttg gcatgtcctg gtgaaacccg cgcactctgc tctgaggcga gtgaggactg  
 23641 tttgtctcag cagcaataaa tcctcccaag agtggatttg gacttttagg aggtgtgatg

8/27

23701 tttggtttat tttcaaatga caagtgcaca gcaccaactg cagcgcttta atctctaaac  
23761 aatgttaaag ctcacaaagg tttctccct cgccctagg atcacctccc aagcttcagg  
23821 gcacaataga ttttttctta aagttccaga gtgcagaaga atttccgtgc cattagcccc  
23881 agagccatgt tgaagtcctt ccttgatttg ctgagatatt ctctccgttt aaagaatcaa  
23941 tactataaac ctgggggttc cagagaccag cattgaaagc acagtttatg tttgtagtaa  
24001 tcaaagggaa tccaagcagc aaccaaataa caaaaggaag tgccagtcac tattacaaaa  
24061 tgtcatcatt ggctgataat gtgatcaaca aattaaattc agtgcacagt cactgggcat  
24121 gagtctgggt ggtgacgtgg gttgtgttaa gccagccaac tacttcttcc cttctctatc  
24181 atcccatgg cttcacagaa tagcctgtat cttggcttac agaagacagg aatgtgatgg  
24241 gttttttttt ttctcttttt ctttttcagc tgcacctgtg gcgtatggaa gttctgagcc  
24301 acagctgtga cttcttccat agctgcagca acactggatc ttaaccacac tgtgccactg  
24361 ccacaatggg aactcccccc tttccttttc ttctgaaca tgatggtttt tgaatgaata  
24421 ggtaaatagc taaaatacca ctacacatct attagaatga cccaaatccc aaacactgac  
24481 aataccagct gctgatgaaa atgtagagcg ataggaaagc ttattcactg ctgctggcaa  
24541 tgcgaaatga tacaaccaat ttgtaagaca gtttggcaga ttcttacaaa actaaacata  
24601 ctcttaccat acaatccagc gatggcacac cttagtattt accctcaaaa gctgaaactt  
24661 ttatgtctac acaaaaactt gcatacagat gtttatagca ggtatttttt tttggggggg  
24721 gtgggttttg gttttttctt ttttgggttt ttttgggcca cacctgcggc atatggaggt  
24781 tccaggcta aggggtgaat cagagctaca gctgcccggc tatgccacag ccacagtaac  
24841 ccagatctg agccacatct gcgacctaca ccacagctca tggccacact ggatccttaa  
24901 cccctgagc gaggccaggg attgaacctg caactagggt gggttggtga ccactgagcc  
24961 acgaaaggaa ctccgtttat agcactttta ttcataattg cccaaaactt ggaagtaaca  
25021 ggtcctttag taggcaaatt aattgataaa ttgtagcaca ttcagacagt ggagtattag  
25081 cagtgtctta aaaaaaaaaa tcaatccatg aaaagacatg gaggaacctc aaatgttcgt  
25141 tacciaaaaa ccatctgcaa agactaccta ctatatggtt ccaactatat gacattctgg  
25201 aaaagggtcaa aggtactgat ccattggagc agtaaaaaata tcagtagatg ccaggcgatg  
25261 gtgtgggaga atggatgaat agacagaaca caggattttt tagggcagtg aaaatactct  
25321 gtatggtacc atacatagtg atagataaat gtcattttac atttggtcag attcatagaa  
25381 agtctgccac caagagtga ccttagtgta aactacgggc ttggtgtgat aatgatgtgt  
25441 cagtgtagtt tcatcagttg caacacgtgc taccctgggt ctggatgttg atggtgggag  
25501 aagctgggag tgggagtggg gggagcgggg ggaattctgc actttcagct caattttgct  
25561 gtgagcctaa acatgttcta aaaactaaaa tctatttttt taaatgaaac gagatagtta  
25621 aagtaagggt gaatcagagg tcgtttgtct tcatatttgg atgggtttta taatgatggt  
25681 gtctggataa tgggctgggt tttgggggtg cagcggaaac tcacagggtta aatcagtaag  
25741 atggaagaat atccatattc cataatccat tgtttgggtc ccagagaagg ataagagagt  
25801 ggcgatgggt aggttaatg acccttgcac atcacgcttg catctggtag gtcagtaaaa  
25861 gtattggaat ataggctggg gtttttgagg aagagctaag aactagtgcc cctgatggt  
25921 aatcaaaggg tacaaaactg cagttataag atgaataagc tctgggatct aatgtcagca  
25981 atgggattgt ttttaatcat actgtaatat atactttaag gttggtcaga gaataagtc  
26041 taagttttct tgccacaaaa aaatggtaat tatgtgatat gatgaagggt tcaagtaatg  
26101 ccacagtggg aatcatttct aatatatgag tgtaccaggg agtttccgtt gtcactcaag  
26161 tgggttaaga atctgactag taaccatgag gatttgggtt cgatccctgg ccttgctcag  
26221 tgggttaaaag gatcaagcat tgccaccagc tgcggtgtag gtcatagaca aagctcagat  
26281 ctggtgttgc tgtggctgtg gcatgggcca gcagctgcaa ctgcaattag acccctagtc  
26341 tgggaacctc catatgctgc ggatgtggcc ctaaaaagcc aaaaaataaa aattaaaaat  
26401 aagtgcacat catcaacact ttgcacacct taaacttaca caatgatata tgtcagttat  
26461 atctgaataa tgcaggaaaa aaggaagaac tagcatctga actgattgat aaggacaagg  
26521 ggataaagac taattaaggc agatggtaat gagttatcac cttcattaca gttaaagctg  
26581 gtgttgaggt atgtggctta taactgcagg ccctgggctt ccaaactctg aaccacagaa  
26641 ttagaaagac ttaggcacac aggcacatac agcacaagta actgcaggcc tctggatgtg  
26701 ggagcagggc ctgtgccact aaggtaccac tgcaaggccc agagcagggt gagcatgccc  
26761 ttctctctgc ctggttggt tccaagagt cagtcaggta agccactcca cctcccaaaa gctgaatgac  
26821 agccaccctg ccagaatcac cagtcaggta agcaacagac gtctctcatc caggatgggg  
26881 tgaatggtg agagtagctg ggaatgttac agcaacagac tgataatagc atattgtctc  
26941 aaaaaatcatt ctttctctaa actgcaaaat acagactaga tatttcag--  
27001 ctctagaaat cccagaggtt acatttaccc cattcttctt tttttcag--

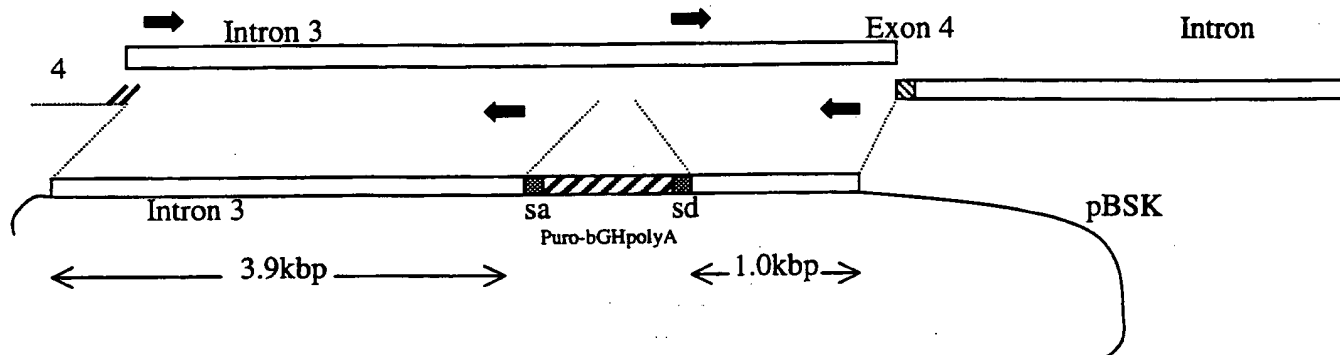
----- --EXON 9

9/27

**Figure 2**

Promoter trap Construct: Insertion of Engineered exon into the intron of a gene.

Endogenous alpha-Gal transferase gene diagram




PCR Primers →

sa = splice acceptor

sd = splice donor

pBSK = pBluescript cloning vector

Puro-bGHpolyA  = puromycin resistance with the bovine Growth hormone gene poly A acceptor signal sequence

10/27

**Figure 3****A.**

10 20 30 40 50  
**ggcggccgcAGGCCTCACTGGCCTAATACGACTCACTATAGGGAGCTCGA**  
CAGCTGAGATCCGGAGTGACCGGATTATGCTGAGTGATATCCCTCGAGCT

60 70 80 90 100  
GGATCAATTAGAGGTCCACCATCCCTTTCCTGAATGCCTAAGGCCAGATA  
CCTAGTTAATCTCCAGGTGGTAGGGAAAGGACTTACGGATTCCGGTCTAT

110 120 130 140 150  
TGTTGGAATTTAGAATTTTTCAAATGCAGAATATTTATCCTATATTATGT  
ACAACCTTAAATCTTAAAAAGTTTACGTCTTATAAATAGGATATAATACA

160 170 180 190 200  
AACGCCCCCAGTGCAGCAACAGCACATAACAATGTACATCAATATTTATG  
TTGCGGGGGTCACGTCGTTGTCGTGATTGTTACATGTAGTTATAAATAC

210 220 230 240 250  
CAAAGAAATGTTTAAACAGTCTCACTAAGTGATAAAGATTATAAATAGCC  
GTTTCTTTACAAATTTGTCAGAGTGATTCACTATTTCTAATATTTATCGG

260 270 280 290 300  
TCGTGTCAGGGCCTGGATGCCAACTGAATTATAAACAGGCTTTTGGTTTT  
AGCACAGTCCCGGACCTACGGTTGACTTAATATTTGTCCGAAAACCAAAA

310 320 330 340 350  
CAGAGCTTGGAGTTGGATGAGGGTCTGAGAACTGCTCCATGTTTCAGGGT  
GTCTCGAACCTCAACCTACTCCCAGACTCTTTGACGAGGTACAAGTCCCA

360 370 380 390 400  
TACCCAGTCTGTGGGTGTCTCCAGACCCACCTCCTTCCCAAGCTCTCTC  
ATGGGTGACACACCCACAGAGGTCTGGGGTGGAGGAAGGGTTCGAGAGAG

410 420 430 440 450  
ACCACCCACACTTCTCTGGGAGTGAAGACAACGGCAGAGAGGCATGGCCA  
TGGTGGGTGTGAAGAGACCCTCACTTCTGTTGCCGTCTCTCCGTACCGGT

460 470 480 490 500  
CAGTGGCCACAGTCTCCACCCCGATCTGTCTGCTCCCAAACCCAGGCCTT  
GTCACCGGTGTCAGAGGTGGGGCTAGACAGACGAGGGTTTGGGTCCGGAA

510 520 530 540 550  
TCCTCGCACTCAGTGCTAATGCTGTTGATGTAGGAGTCAAGTGGCTTTTT  
AGGAGCGTGAGTCACGATTACGACAACCTACCTCAGTTCACCGAAAAA

560 570 580 590 600  
CCAGCATCTGGGCCGAGCTGCATGTAGCCCCGTGCATTTTCGTAACCTTGC

71/27

GGTCGTAGACCCGGCTCGACGTACATCGGGGCACGTAAAGCATTGAAACG

610 620 630 640 650  
CCTGGGCCCCGGGCTGTTTGTGCCAGGACCTGAGGTGCTCACAGGAAAGA  
GGACCCGGGGCCCGACAAACACGGTCCTGGACTCCACGAGTGTCTTTCT

660 670 680 690 700  
AGCTCCATCTCCCCATCTTCTCACCATCTCTGGAACACCACCTATCATGA  
TCGAGGTAGAGGGGTAGAAGAGTGGTAGAGACCTTGTGGTGGATAGTACT

710 720 730 740 750  
TTGTATCTGAAAGGTGGCGATTGAATCAGAGGTTTCCAAACAGAGCTCAC  
AACATAGACTTTCCACCGCTAACTTAGTCTCCAAAGGTTTGTCTCGAGTG

760 770 780 790 800  
TTTCCAAGCAAGAAGGAATAGAGTGACATGGCTGATAATCCCATACTTTC  
AAAGGTTTCGTTCTTCTTATCTCACTGTACCGACTATTAGGGTATGAAAG

810 820 830 840 850  
TCTTCTTTAACTGGATTTTCAACAAGAGGTGATGGAGCACCTGAGATCTA  
AGAAGAAATTGACCTAAAGTGTGTCTCCACTACCTCGTGGACTCTAGAT

860 870 880 890 900  
AGCCTGGAGTCACCTCAGAACCCTCTCTGCAAATATTTGGAGAATAACCC  
TCGGACCTCAGTGGAGTCTTGGGAGAGACGTTTATAAACCTCTTATTGGG

910 920 930 940 950  
GTCCCCTGAAAGGACACATCTCAGTGCCATTCTCATTTTCATTCACACATC  
CAGGGGACTTTCCTGTGTAGAGTCACGGTAAGAGTAAAGTAAGTGTGTAG

960 970 980 990 1000  
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGCTTTTTTGCCATTTCTTGGG  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAACCCGAAAACGGTAAAGAACCC

1010 1020 1030 1040 1050  
CCAGTCCTGCGGCATATGGAGGTTCCCAGGCTAAGGGTCTAATTGGAGCC  
GGTCAGGACGCCGTATACCTCCAAGGGTCCGATTCCCAGATTAACCTCGG

1060 1070 1080 1090 1100  
GTAGCTGCAGGCCTACGCCAGAGCCAGAGCCACACGGGATCTGAGCCGCG  
CATCGACGTCCGGATGCGGTCTCGGTCTCGGTGTGCCCTAGACTCGGCGC

1110 1120 1130 1140 1150  
TCTGCAACCTACACCACAGCTCACGGCAACGCCGGATCCTTAAGCCACTG  
AGACGTTGGATGTGGTGTGAGTGCCGTTGCGGCCTAGGAATTTCGGTGAC

1160 1170 1180 1190 1200  
AGCAAGGCCAGGGATGGAACCCACAACCTCATGTTTCCTAGTCAGATTTCG  
TCGTTCCGGTCCCTACCTTGGGTGTTGGAGTACAAAGGATCAGTCTAAGC

1210 1220 1230 1240 1250  
TTAACCACAGAGCCACAACGGGAACCTCCACACATTATTTATTGACGGCC  
AATTGGTGTCTCGGTGTTGCCCTTGAGGGTGTGTAATAATAACTGCCGG

12/27

1260 1270 1280 1290 1300  
TTCTCTGCTCTCTGTGGGGCACTGGGAATTCAGGGGTGATCAAGAAGTCA  
AAGAGACGAGAGACACCCCGTGACCCTTAAGTCCCCACTAGTTCTTCAGT

1310 1320 1330 1340 1350  
TCCCTCCTGCCCTCAGGAAGCTCAAACCACTCATTATTTATTGACGGCCT  
AGGGAGGACGGGAGTCCTTCGAGTTTGGTGAGTAATAAATAACTGCCGGA

1360 1370 1380 1390 1400  
TCTCTGCTCTCTGTGGGGCACTGGGAATTCAGGGGTGACGAAGAAGTCAT  
AGAGACGAGAGACACCCCGTGACCCTTAAGTCCCCACTGCTTCTTCAGTA

1410 1420 1430 1440 1450  
CCCTCCTGCCCTCAGGAAGCTCAAACAAGCAGGTAGAGGAGGCAGAGCAA  
GGGAGGACGGGAGTCCTTCGAGTTTGTTCGTCCATCTCCTCCGTCTCGTT

1460 1470 1480 1490 1500  
AATGCAGGTCTTATCCGGTGAGCCGACTCCCAGGGCGATGTGTACAGCAA  
TTACGTCCAGAATAGGCCACTCGGCTGAGGGTCCCGCTACACATGTCGTT

1510 1520 1530 1540 1550  
AGGAATAGAGGGATGGGGGCCGGAGGAGAGAAAAGGGCTTCAGCCGTGGT  
TCCTTATCTCCCTACCCCCGGCCTCCTCTCTTTTCCCGAAGTCGGCACCA

1560 1570 1580 1590 1600  
CAGGGTGGGGGTGGGAAGTGGCTTCACAAAGGCAGTGACATTGGCTCCCA  
GTCCCACCCCCACCCTTCACCGAAGTGTTTCCGTCACTGTAACCGAGGGT

1610 1620 1630 1640 1650  
GGTGTCCACTCTTCTGTCTCTGTCTACCTTCTGGTCCTCTCCTTCTGGGCC  
CCACAGGTGAGAAGACAGAGACGATGGAAGACCAGGAGAGGAAGACCCGG

1660 1670 1680 1690 1700  
CTCCTCTATCCTACCTCTAAAGCTTCACCCACATCCTCCTTTCTTTTCT  
GAGGAGATAGGATGGAGATTTCGAAGTGGGTGTAGGAGGAAAGGAAAAGA

1710 1720 1730 1740 1750  
CTCTCTGGATTCTCTCCTGGGTAATCAAATTCGTTCCCTTCACGTCAGAT  
GAGAGACCTAAGAGAGGACCCATTAGTTTAAGCAAGGGAAGTGCAGTCTA

1760 1770 1780 1790 1800  
CCGATACGTTCCCTTGGTCCATGAACAACCTTCTCCGATTGCATGGTCTGCC  
GGCTATGCAAGGAACCAGGTACTTGTTGAAGAGGCTAACGTACCAGACGG

1810 1820 1830 1840 1850  
TACATCTCTCTGATGAACCTTAGACTTGAATGTCCACTTGTCTCCCTGTCT  
ATGTAGAGAGACTACTTGAAATCTGAACTTACAGGTGAACAGAGGGACAG

1860 1870 1880 1890 1900  
CCCTTTTAGGTATTTCGCACACTCCCCGACATTCACACGTCCAAAAGGGAA  
GGGAAAATCCATAAGCGTGTGAGGGGCTGTAAGTGTGCAGGTTTTCCTT

1910 1920 1930 1940 1950  
TTCATGATTATTATCCTCCAAGCCTGTTCTCCTCCAGCCCATCTGAGAA

13/27

AAGTACTAATAATAGGAGGTTCCGACAAGGAGGAGGTCGGGTAGACTCTT

1960 1970 1980 1990 2000  
AATACTACAACCCCCCTGCTTAAGCAGAAATCTTGGGTCTTCCTTGTCTC  
TTATGATGTTGGGGGACGAATTCGTCTTTAGAACCCAGAAGGAACAGAG

2010 2020 2030 2040 2050  
ATCTCTGATAACAAAATTACCAACCACGTCCTATCAATTCTCTCTCCAAA  
TAGAGACTATTGTTTTAATGGTTGGTGCAGGATAGTTAAGAGAGAGGTTT

2060 2070 2080 2090 2100  
GTATATATATATATATATATTTTTTTAATTTTTTCCCGCTGTACAGCATGGG  
CATATATATATATATATAAAAAAATTAAAAAGGGCGACATGTCGTACCC

2110 2120 2130 2140 2150  
GATCAAGTTATTCTTACATGTATATTTTTCCCCCACCCTTTGTTCCGTTG  
CTAGTTCAATAAGAATGTACATATAAAAGGGGGGTGGGAAACAAGGCAAC

2160 2170 2180 2190 2200  
CAATATGAGTATCTAGACATAGTTCTCAATGCTACTCAGCAGGATCTCCT  
GTTATACTCATAGATCTGTATCAAGAGTTACGATGAGTCGTCCTAGAGGA

2210 2220 2230 2240 2250  
TGTAATATAAGTTGTATCTGATAACCCCAAGCTCCCGATCCCTCCCACT  
ACATTTATATTCAACATAGACTATTGGGGTTCGAGGGCTAGGGAGGGTGA

2260 2270 2280 2290 2300  
CCCTCCCTCTCCTGTCGGGCAGCCACAAGTCTATTCTCCAAGTCCATGAT  
GGGAGGGAGAGGACAGCCCGTCGGTGTTCAGATAAGAGGTTTCAGGTACTA

2310 2320 2330 2340 2350  
TTTCTTTTCTGTGGAGATGGTCATTTGTGCTGGATATTAGATTCCAGTTA  
AAAGAAAAGACACCTCTACCAGTAAACACGACCTATAATCTAAGGTCAAT

2360 2370 2380 2390 2400  
TAAGTGATATCATATGGTATTTGTCAAAGTATATATTTTATTTTCTTTG  
ATTCATATAGTATACCATAAACAGTTTCATATATAAAATAAAAAGAAAC

2410 2420 2430 2440 2450  
TCTTTTGTCTTTTGTCTTTTTTTTGTGTTGTTGTTGTTGTTGTTGTTG  
AGAAAAACAGAAAACAGAAAAAACAACAACAACAACAACAACAAC

2460 2470 2480 2490 2500  
TTGTTGCTATTACTTGGGCGCTCCCGCGGCATATGGAGGTTCCCAGGCT  
AACAACGATAATGAACCCGCGAGGGCGCCGTATACCTCCAAGGGTCCGA

2510 2520 2530 2540 2550  
AGGAGTTGAATCGGAGCTGTAGCCACCGGCCTACGCCAGAGCCACAGCAA  
TCCTCAACTTAGCCTCGACATCGGTGGCCGGATGCGGTCTCGGTGTCGTT

2560 2570 2580 2590 2600  
CGCGGGATCCGAGCCGCGTCTGCAACCTACACCACAGCTCACGGCAACGC  
GCGCCCTAGGCTCGGCGCAGACGTTGGATGTGGTGTGAGTGCCGTTGCG



14/27

2610 2620 2630 2640 2650  
TGGATCCTTAACCCACTGAGCAAGGGCAGGGACCGAACCCGCAACCTCAT  
ACCTAGGAATTGGGTGACTCGTTCCCGTCCCTGGCTTGGGCGTTGGAGTA

2660 2670 2680 2690 2700  
GGTTCCTAGTCGGATTTCGTTAACCCTGCGCCACGACGGGAACTCCCAA  
CCAAGGATCAGCCTAAGCAATTGGTGACGCGGTGCTGCCCTTGAGGGTTT

2710 2720 2730 2740 2750  
GTATATTTTGAATCAAGCCACCCTTTGAGCCAGGCCACCTCCTCTTTATG  
CATATAAACTTAGTTTCGGTGGGAACTCGGTCCGGTGGAGGAGAAATAC

2760 2770 2780 2790 2800  
GTCATGAGAACGGTCTGCCCTTGTCTTTTCTCCATTCTCCACACTCAGC  
CAGTACTCTTGCCAGACGGGAACAGGAAAAGAGGTAAGAGGTGTGAGTCG

2810 2820 2830 2840 2850  
ACCCAGATGGGTCTCTCTAGGTGAAGTTGGATCAGGGGATTCTCCAGCTT  
TGGGTCTACCCAGAGAGATCCACTTCAACCTAGTCCCCTAAGAGGTCGAA

2860 2870 2880 2890 2900  
TAGATGCTTTTTGGGATTCCCCACCCTACTTTCCATACCTTTCCAGGTTC  
ATCTACGAAAAACCTAAGGGGTGGGATGAAAGGTATGGAAAGGTCCAAG

2910 2920 2930 2940 2950  
TGACTGCCTCTGCCCCCTTCTGACTGCCTAGCACCAGCCACTCAAGGGG  
ACTGACGGAGACGGGGGGAAGACTGACGGATCGTGGTCGGTGAGTTCCCC

2960 2970 2980 2990 3000  
GACAGTGTCACTCACTATTTTTTTTCTTGTCCAGGTTTTTTTGCTTTTGT  
CTGTCAAGTCACTGATAAAAAAAGAACAGGTCCAAAAACGAAAACAAA

3010 3020 3030 3040 3050  
TTTTCAAACACGAGCAGCTCTTTCTCTTGTCTGCCTGGTATAGATGCTGT  
AAAAGTTTGTGCTCGTCGAGAAAGAGAACAGACGGACCATATCTACGACA

3060 3070 3080 3090 3100  
TTCCAAAATATTCTCATCCCTTCTCACGGCCCTTGTATCCTTTCCCATC  
AAGGTTTTATAAGAGTAGGGAAGAGTGCCGGGAACAGTAGGAAAGGGTAG

3110 3120 3130 3140 3150  
CTATCTTCATCCCTTGGAAGCTCTAAAGTCATCTCCCCAAATTGAAGGG  
GATAGAAGTAGGGAACCCTTCGAGATTTTCAGTAGAGGGGTTTAACTTCCC

3160 3170 3180 3190 3200  
TGACTAAAGAGTTTCCCAGAAGGAAAACTGAGTTTCCAACACTACTACT  
ACTGATTTCTCAAAGGGTCTTCCTTTTGGACTCAAAGGTGATGATGTGA

3210 3220 3230 3240 3250  
GACTTGCAAGAAATGTTTGTGTCTTCATTAAATGAAAAAGAAAAACTGT  
CTGAACGTTCTTTACAAACACAGAAGTAATTTACTTTTTCTTTTTTGACA

3260 3270 3280 3290 3300  
AACAAGATATGAGAAAATACAGAAAGGAAATAATAAGACTAGAAAAGTCA

15/27

TTGTTCTATACTCTTTTATGTCTTTCCTTTATTATTCTGATCTTTTCAGT

3310 3320 3330 3340 3350  
AATATATAGTGAAGGTGTTGCATCAAACACTTAAATAAACTAGTACAGAT  
TTATATATCACTTCCACAACGTAGTTTGTGAATTTATTTGATCATGTCTA

3360 3370 3380 3390 3400  
GTTAAAAGACTAAATTATATAGTTGAAGGATAGCTGTGAAGATGTAAACT  
CAATTTTCTGATTTAATATATCAACTTCCTATCGACACTTCTACATTTGA

3410 3420 3430 3440 3450  
ATGACATCTAAAACACAAAATGTTGGCGTTCCCGTCACGGCACAGTGGA  
TACTGTAGATTTTGTGTTTTTACAACCGCAAGGGCAGTGCCGTGTCACCTT

3460 3470 3480 3490 3500  
ACGAATCCGACTAGGAACCATGAGGTTGCAGGTTCAATTCCTGCCCTTGC  
TGCTTAGGCTGATCCTTGGTACTCCAACGTCCAAGTTAAGGACGGGAACG

3510 3520 3530 3540 3550  
TCAGTGGGTAAAGGATCCGGTGTTGCCGTGAGCTGTGGTGTAGGTAGCCA  
AGTCACCCAATTCTAGGCCACAACGGCACTCGACACCACATCCATCGGT

3560 3570 3580 3590 3600  
ATGAGGCTTGGATCCCGCGTTGCTGTGGCTCTGGTGTAGGCCGGTGGCTA  
TACTCCGAACCTAGGGCGCAACGACACCGAGACCACATCCGGCCACCGAT

3610 3620 3630 3640 3650  
CAGCTCCGATTTCGACCCCTAGCCTGGGAACCTCCATATGCCGCGGGAGCG  
GTCGAGGCTAAGCTGGGGATCGGACCCCTGGAGGTATACGGCGCCCTCGC

3660 3670 3680 3690 3700  
GGCCCTTAAAAAGACAAAAGACCAAAAAAAAAAAAAAAAAACAAAAACCCAC  
CCGGGAATTTTTCTGTTTTCTGGTTTTTTTTTTTTTTTGTTTTTTGGGTG

3710 3720 3730 3740 3750  
AAAATGTTGGGAATCAGTCCTCTACTAGTATTATGTTATTGTCAAGTTTT  
TTTTACAACCCTTAGTCAGGAGATGATCATAATAACAATAACAGTTCAAAA

3760 3770 3780 3790 3800  
CCTTTTATGTCTGTTAATATTTGCGTTCTAGATGTAGGTGCTCTGATATC  
GGAAAATACAGACAATTATAAACGCAAGATCTACATCCACGAGACTATAG

3810 3820 3830 3840 3850  
GTGTGCATATATGTTAACCAATGTTATGTCTTCCTCTGGTATTGATCCCT  
CACACGTATATACAATTGGTTACAATACAGAAGGAGACCATAACTAGGGA

3860 3870 3880 3890 3900  
TTGTTATTATGTAATGCCCTACTTTATCTTTTGTACATTCTTTGTTTAT  
AACAATAATACATTACGGGATGAAATAGAAAACAATGTAAGAAACAATA

3910 3920 3930 3940 3950  
GAGTATTGCTGATATGTGGCTAGCTGCCACACTTTTCTTGTCCTTTCCAT  
CTCATAACGACTATACACCGATCGACGGTGTGAAAAGAACAGGAAAGGTA

16/27

3960 3970 3980 3990 4000  
 TTACAATAAATATCTTTCTATCTCCACCCAAATTAAAGTACTCCGCAACC  
 AATGTTATTTATAGAAAGATAGAGGTGGGTTTAATTTTCATGAGGCGTT**GG**

4010 4020  
 TGTTATTCCACCCAGCATCCgtcgcac  
**ACAATAAGGTGGGTCTAGGcagctg**

**B.**

Ligated to the following intron 4 splice acceptor sequence

Intron 4 splice acceptor sequence

11530 11540 11550  
gtcgcac**CCACCGTTTGATCTGAGTAATTCTGAAATG**  
cagctgGGTGGCAAACCTAGACTCATTAAAGACTTTAC

11560 11570 11580 11590 11600  
 ACGAGAGTCCCGTGATATCATTTTTTCGATCTCGAAGGTGGAAACCTGGG  
 TGCTCTCAGGGCACTATAGTAAAAAGCTAGAGCTTCCACCTTTGGACCC

11610 11620 11630 11640 11650  
 AGTAGCCACAACCCAGGCTCTCAGCTCAGCCTAGGGTTTCAATGATAATG  
 TCATCGGTGTTGGGTCCGAGAGTCGAGTCGGATCCCAAAGTTACTATTAC

11660 11670 11680 11690 11700  
 ATTGCAAAATAGCTTTTCTCTGCATTCCAAGTAACATGATATGTTTTTAT  
 TAACGTTTTATCGAAAAGAGACGTAAGGTTTATTGTACTATACAAAATA

11710  
 TTCCATTTGCTTTTAGgaattc  
**AAGGTAAACGAAAATCcttaag**

**C.**

Ligated to Neomycin Resistance gene

gaattc**AATGGATCCCCACCATGG-NEO-GGATATCCACTACTTAGTAATAGCCG**aaagctt  
**CCTATAGGTGATGAATCATTATCGGcttcgaa**

17/27

**D.**

Ligated to the intron 4 splice donor sequence

4940                      4950  
aagcttGTAATTATGAAAC  
ttcgaaCATTAATACTTTG

4960                      4970                      4980                      4990                      5000  
**ATGATGAAATGATGTTGATGAAAGTCTCCTCTAATCTCCTAGTTATCAGC**  
**TACTACTTTACTACAACACTACTTTCAGAGGAGATTAGAGGATCAATAGTCG**

5010                      5020                      5030                      5040                      5050  
**CAAGTCACCAGCTTGCAATTAAGTAGGATTCACTGACACCGTAAAGAAA**  
**GTTCAAGTGGTCGAACGTAATTTTCATCCTAAGTGAAGTGTGGCATTTCCTT**

5060                      5070                      5080                      5090                      5100  
**GCATTCCAGAGAGTTGCCGTTGTGGCTCAGGGGCAGCAAACCCAATTAGG**  
**CGTAAGGTCTCTCAACGGCAACACCGAGTCCCCGTCGTTTGGGTTAATCC**

5110                      5120                      5130                      5140                      5150  
**ATCCAAGAGGAGGTGGGTTTGATCCCTGGCCTTGCTCTTTGGCTTAAGGA**  
**TAGGTTCTCCTCCACCCAACTAGGGACCGGAACGAGAAACCGAATTCCT**

5160                      5170  
**TCCGGCATTGCCGTGACCTGTGG**ctgcag  
**AGGCCGTAACGGCACTGGACACC**gacgtc

**E.**

Ligated to intron 3 3' sequence

ctgcag**CCCTCTTCAACTACAATTTTCATGCAGC**  
GGGAGAAGTTGATGTTAAAGTACGTCG

4060                      4070                      4080                      4090                      4100  
**GATCAAGAAATAGAATGTACCGACTGTTTGCCATTGGTGGGGCATGGGAA**  
**CTAGTTCTTTATCTTACATGGCTGACAAACGGTAACCACCCCGTACCCCTT**

4110                      4120                      4130                      4140                      4150  
**AAGTGGGTGGAAAGTGCAGAGCTTAGATTATAAAGGCCAGGGTGAGAGTT**  
**TTACCCACCTTTCACGTCTCGAATCTAATATTTCCGGTCCCCTCTCAA**

4160                      4170                      4180                      4190                      4200  
**CCCATTGTGGTCAGCTGAAATGAATCTGACTAGCATCCATGAGCACGAAG**  
**GGGTAACACCAGTCGACTTTACTTAGACTGATCGTAGGTACTCGTGCTTC**

4210                      4220                      4230                      4240                      4250

18 / 27

GTTTGATCCCTGGCCTCAATCAGTGGGTTAAGGATCTGGCGTTGCTGTCC  
CAAAC TAGGGACCGGAGTTAGTCACCCAATTCCTAGACCGCAACGACAGG

4260 4270 4280 4290 4300  
GTGAGTTGTGGTGTAGTTCGCAGACAAGGCGTGGACTTAGTGTGGCTGTG  
CACTCAACACCACATCAAGCGTCTGTTCCGCACCTGAATCACACCGACAC

4310 4320 4330 4340 4350  
GCTGTGGCATAGGCTAGTGGCTACAGCTCTGATTCGACCCCTAGCCTGGG  
CGACACCGTATCCGATCACCGATGTGCGAGACTAAGCTGGGGATCGGACCC

4360 4370 4380 4390 4400  
AATCTCTATATGCTGTGAGTGTGGCCCTAAAATTTAAATGAAATTAAATA  
TTAGAGATATACGACACTCACACCGGGATTTTAAATTTACTTTAATTTAT

4410 4420 4430 4440 4450  
AAGGACCAGGGTATATTTTTCTTTGAGGATAAGGTACATAGTCAGTATAT  
TTCTTGGTCCCATATAAAAAGAACTCCTATTCCATGTATCAGTCATATA

4460 4470 4480 4490 4500  
CAAGGACAGTAGACCTAGGAAACGGATGCTTCCTCTAGTCTGTGATGCGA  
GTTCTGTCACTCTGGATCCTTTGCCTACGAAGGAGATCAGACACTACGCT

4510 4520 4530 4540 4550  
GGTGGGGCATCTGAGTTGGGGGCGGCTGGAGCCCTTAGGGACCATTAACT  
CCACCCCGTAGACTCAACCCCGCCGACCTCGGGAATCCCTGGTAATTGA

4560 4570 4580 4590 4600  
AAACCCGTCACCTCTCCACATCTCGGTGGACCTTGGGATCAGTCAGGATG  
TTTGGGCAGTGAGAGGGTGTAGAGCCACCTGGAACCCTAGTCAGTCCTAC

4610 4620 4630 4640 4650  
CTTCCCCTTTGAGCCTCAAAATGGCCTTAGTATCCTTCCCAACCCAGACG  
GAAGGGGAACTCGGAGTTTACC GGAATCATAGGAAGGGTTGGGTCTGC

4660 4670 4680 4690 4700  
GCCCTGTCAGTTCATTGACTTGGCTAATTTGCCAGTGTAGGCCTATGCAA  
CGGGACAGTCAAGTAACTGAACCGATTAAACGGTCACATCCGGATACGTT

4710 4720 4730 4740 4750  
ATTAAGGTAGAACGCACTCCTTAGCGCTCGTTGACTATTCATCAACTTTT  
TAATTCCATCTTGCGTGAGGAATCGCGAGCAACTGATAAGTAGTTGAAAA

4760 4770 4780 4790 4800  
CCTTTTAGAAAAGATATTGGTATAAGCACTTCTTAAAAAACCATATTCCA  
GGAAAATCTTTTCTATAACCATATTCGTGAAGAATTTTTTGGTATAAGGT

4810 4820  
CTCTGGGTGTATTTAATCTAATTTTCctcgag  
**GAGACCCACATAAATTAGATTAAAAAGgagctc**

19/27

**Figure 4**

taataatgga tccccacc**AT** GGGCATCGAG CAGGACGGCC TGCACGCCGG CAGCCCCGCC  
 GCCTGGGTGG AGAGACTGTT CGGCTACGAC TGGGCCCAGC AGACCATCGG CTGCAGCGAC  
 GCCGCCGTGT TCAGACTGAG CGCCCAGGGC AGACCCGTGC TGTTCGTGAA GACCGACCTG  
 AGCGGCGCCC TGAACGAGCT GCAGGACGAG GCCGCCAGAC TGAGCTGGCT GGCCACCACC  
 GGCCTGCCCT GCGCCGCCGT GCTGGACGTG GTGACCGAGG CCGGCAGAGA CTGGCTGCTG  
 CTGGGCGAGG TGCCCGGCCA GGACCTGCTG AGCAGCCACC TGGCCCCCGC CGAGAAGGTG  
 AGCATCATGG CCGACGCCAT GAGAAGACTG CACACCCTGG ACCCCGCCAC CTGCCCCCTC  
 GACCACCAGG CCAAGCACAG AATCGAGAGA GCCAGAACCA GAATGGAGGC CGGCCTGGTG  
 GACCAGGACG ACCTGGACGA GGAGCACCAG GGCCTGGCCC CCGCCGAGCT GTTCGCCAGA  
 CTGAAGGCCA GAATGCCCGA CGGCGAGGAC CTGGTGGTGA CCCACGGCGA CGCCTGCCTG  
 CCCAACATCA TGGTGGAGAA CGGCAGATTG AGCGGCTTCA TCGACTGCGG CAGACTGGGC  
 GTGGCCGACA GATACCAGGA CATCGCCCTG GCCACCAGAG ACATCGCCGA GGAGCTGGGC  
 GGCAGTGGG CCGACAGATT CCTGGTGCTG TACGGCATCG CCGCCCCCGA CAGCCAGAGA  
 ATCGCCTTCT ACAGACTGCT GGACGAGTTC TTCTGAATCT TGCAGCTGGT GGGGTTCTGG  
 ATATCCACTA <sup>\*</sup>CTTAGccgaa ttctatatat

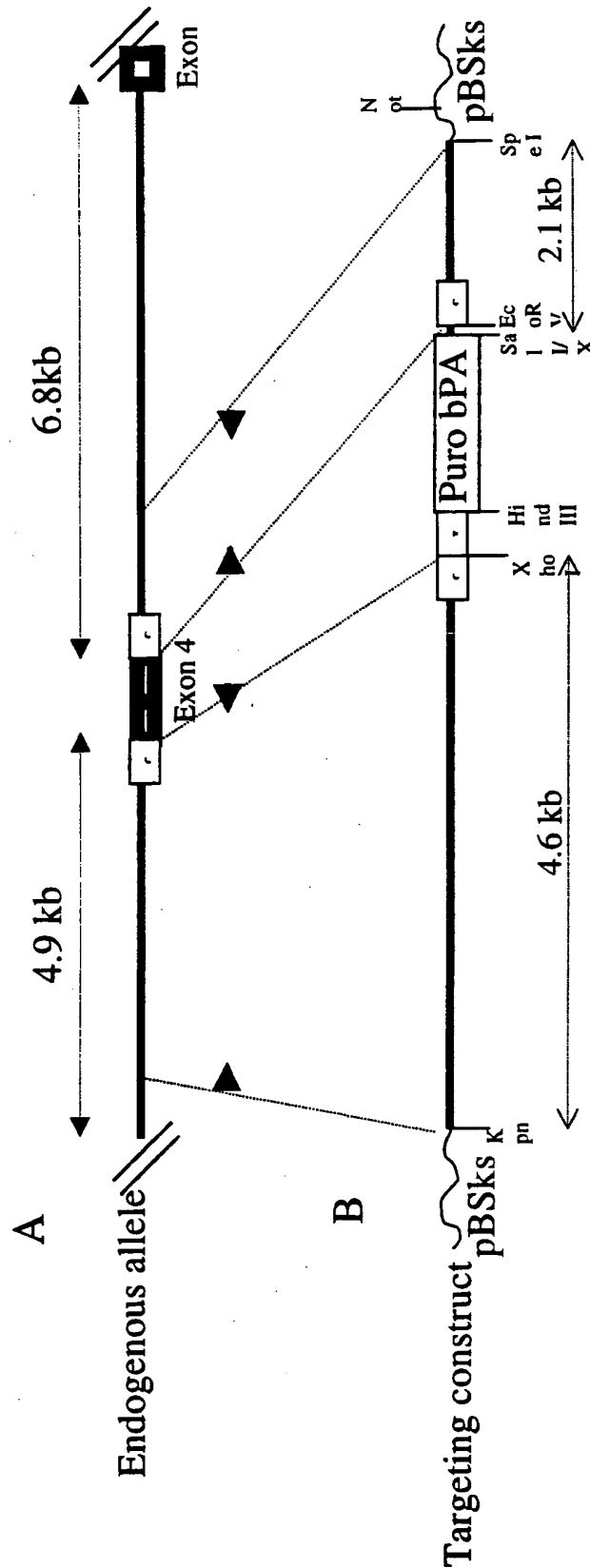
20/27

**Figure 5**

```
1 ctgaagctta ccatgaccga gtacaagccc acggtgcgcc tcgccacccg cgacgacgtc
61 ccccgggccg tacgcaccct cgccgccgcg ttcgccgact accccgccac gcgccacacc
121 gtcgaccccg accgccacat cgagcgggtc accgagctgc aagaactctt cctcacgcgc
181 gtcgggctcg acatcgga aaagtgtgggtc gcggacgacg gcgccgcggt ggcggtctgg
241 accacgcccg agagcgtcga agcgggggcg gtgttcgccg agatcggccc gcgcatggcc
301 gagttgagcg gttcccggct ggccgcgcag caacagatgg aaggcctcct ggcgccgcac
361 cggcccaagg agcccgcgtg gttcctggcc accgtcggcg tctcgcccga ccaccagggc
421 aagggtcttg gcagcgcgt cgtgctcccc ggagtggagg cggccgagcg cgcgggggtg
481 cccgccttcc tggagacctc cgcgccccgc aacctcccct tctacgagcg gctcggcttc
541 accgtcaccg ccgacgtcga ggtgcccga ggaccgcga cctggtgcat gaccgcaag
601 cccggtgcct gacgcccgc ccacgaccg cagcgcgcga ccgaaaggag cgcacgaccc
661 catgcatcga tgatctagag ctcggtgatc agcctcgact gtgccttcta gttgccagcc
721 atctgttgtt tgcccctccc ccgtgccttc cttgaccctg gaaggtgcca ctcccactgt
781 cctttcctaa taaaatgagg aaattgcac gcattgtctg agtaggtgtc attctattct
841 ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata gcaggcatgc
901 tggggatgcg gtgggctcta tggcttctga ggcggaaaga accagcctcg a
```

21/27

## Figure 6



- ▲: PCR primer positions

**SA: splice acceptor, AG dinucleotide**

SD: splice donor, GT dinucleotide

**KCS: Kozak consensus sequence**

**Puro bpA: puromycin gene coding sequence and**

bovine growth hormone poly A signal sequence

**pBS: Bluescript cloning vector**



22/27

**Figure 7**

```

1  gtcgactcta ggcctcactg gcctaatacag actcactata gggagctcga ggatcaatta
61  gaggtccacc atccctttcc tgaatgccta aggccagata tgttggaatt tagaattttt
121 caaatgcaga atatttatcc tatattatgt aacgccccca gtgcagcaac agcacataac
181 aatgtacatc aatatttatg caaagaaatg tttaaacagt ctcactaagt gataaagatt
241 ataaatagcc tcgtgtcagg gcctggatgc caactgaatt ataaacaggc ttttggtttt
301 cagagcttgg agttggatga gggctctgaga aactgctcca tgttcagggt taccagctct
361 gtgggtgtct ccagacccca cctccttccc aagctctctc accaccaca cttctctggg
421 agtgaagaca acggcagaga ggcattggcca cagtggccac agtctccacc ccgatctgtc
481 tgctcccaaa cccaggcctt tcctcgcact cagtgtcaat gctgttgatg taggagtcaa
541 gtggcttttt ccagcatctg ggcgagctg catgtagccc cgtgcatttc gtaactttgc
601 cctgggcccc gggctgtttg tgccaggacc tgaggtgtct acaggaaaga agctccatct
661 ccccatcttc tcaccatctc tggaaacacca cctatcatga ttgtatctga aaggtggcga
721 ttgaatcaga ggtttccaaa cagagctcac tttccaagca agaaggaata gagtgcacatg
781 gctgataatc ccatactttc tcttctttaa ctggatttca caacagagggt gatggagcac
841 ctgagatcta agcctggagt cacctcagaa ccctctctgc aaatatattgg agaataaccc
901 gtcccttgaa aggacacatc tcagtgccat tctcatttca ttcacacatc tttttttttt
961 tttttttttt ttttttttgg gcttttttgc atttcttggg ccagtcctgc ggcataatgga
1021 ggttcccagg ctaaggggtct aattggagcc gtagctgcag gcctacgcca gagccagagc
1081 cacacgggat ctgagccgcg tctgcaacct acaccacagc tcacggcaac cccggatcct
1141 taagccactg agcaaggcca gggatggaac ccacaacctc atgtttccta gtcagattcg
1201 ttaaccacag agccacaacg ggaactccca cacattatct attgacggcc ttctctgtct
1261 tctgtggggc actgggaatt caggggtgat caagaagtca tccctcctgc cctcaggaag
1321 ctcaaaccac tcattattta ttgacggcct tctctgtctc ctgtggggca ctgggaattc
1381 aggggtgacg aagaagtcat ccctcctgcc ctgaggaagc tcaaacaagc aggtagagga
1441 ggcagagcaa aatgcaggtc ttatccgggt agccgactcc cagggcgatg tgtacagcaa
1501 aggaatagag ggtggggggc cggaggagag aaaagggcct cagccgtggt caggggtggg
1561 gtgggaagtg gcttcacaaa ggcagtgaca ttgggtccca ggtgtccact cttctgtctc
1621 tgctaccttc tggtcctctc cttctggggc ctcctctatc ctacctctaa agcttcaccc
1681 acatcctcct ttccttttct ctctctggat tctctcctgg gtaatcaaat tegtccctt
1741 cacgtcagat ccgatacgtt ccttgggtcca tgaacaactt ctccgattgc atggtctgcc
1801 tacatctctc tgatgaactt tagacttgaa tctccacttg tctccctgtc ccttttagg
1861 tattegcaca ctccccgaca ttcacacgtc caaaagggaa ttcattgatta ttatcctcca
1921 agcctgttcc tctccagcc catctgagaa aatactacaa cccccctgct taagcagaaa
1981 tcttgggtct tcttgtctc atctctgata acaaaattac caaccacgtc ctatcaattc
2041 tctctccaaa gtatatatat atatatattt ttttaatttt ttcccgtgt acagcatggg
2101 gatcaagtta ttcttacatg tatattttcc ccccaccctt tgttccgttg caatatgagt
2161 atctagacat agttctcaat gctactcagc aggatctcct tgtaaatata agttgtatct
2221 gataacccca agtcccgat ccctccact ccctccctct cctgtcgggc agccacaagt
2281 ctattctcca agtccatgat tttcttttct gtggagatgg tcatttgtgc tggatattag
2341 attccagtta taagtatat catatggtat ttgtcaaagt atatatttta ttttctttg
2401 tctttttgtc ttttgtcttt ttttgttgt tgttgttgtt gttgttgttg ttgtgtctat
2461 tacttggggc gctcccgagg catatggagg tccccaggct aggagttaa tcggagctgt
2521 agccaccggc ctacgccaga gccacagcaa cgcgggatcc gagccgctc tgcaacctac
2581 accacagctc acggcaacgc tggatcctta acccactgag caagggcagg gaccgaaccc
2641 gcaacctcat ggttcctagt cggattcgtt aaccactgcg ccacgacggg aactccaaa
2701 gtatatattg aatcaagcca ccctttgagc caggccacct cctctttatg gtcattgagaa
2761 cggctctgcc ttgtcctttt ctccattctc cacactcagc acccagatgg gtctctctag
2821 gtgaagttag atcaggggat tctccagctt tagatgcttt ttgggattcc ccaccctact
2881 ttccatacct ttccagggtc tgactgcctc tgccccctt ctgactgcct caggtttttt gcttttgttt
2941 actcaagggg gacagtgtca gtcactatct ttttcttgtc tagatgctgt ttccaaaata
3001 ttttcaaaaca cgagcagctc tttctcttgt ctgcctggta ctatcttcat cccttgggaa
3061 ttctcatccc ttctcacggc ccttgtcatc ctttcccatc gtttccaga aggaaaaact
3121 gctctaaagt catctcccca aattgaaggg tgactaaaga gtttccaga aatgaaaaag
3181 gagtttccaa ctactacact gacttgcaag aaatgtttgt gtcttcatta aatgaaaaag

```

23/27

```

3241 aaaaaactgt aacaagatat gagaaaatac agaaaggaaa taataagact agaaaagtca
3301 aatatatagt gaaggtgttg catcaaacac ttaaataaac tagtacagat gttaaaagac
3361 taaattatat agttgaagga tagctgtgaa gatgtaaact atgacatcta aaacacaaaa
3421 tgttggcggt cccgtcacgg cacagtggaa acgaatccga ctaggaacca tgaggttgca
3481 ggttcaattc ctgcccttgc tcagtgggtt aaggatccgg tgttgccgtg agctgtgggtg
3541 taggtagcca atgaggcttg gatcccgctg tgctgtggct ctggtgtagg cgggtggcta
3601 cagctccgat tcgaccccta gcctgggaac ctccatattg cgggggagcg ggccttaaa
3661 aagacaaaag accaaaaaaa aaaaaaaca aaaaaccac aaaatgttgg gaatcagtc
3721 tctactagta ttatgttatt gtcaagtttt ccttttatgt ctggttaata ttgcttcta
3781 gatgtagggt ctctgatatc gtgtgcatat atgttaacca atgttatgtc ttcctctggt
3841 attgatccct ttgttattat gtaatgccct actttatctt ttgttacatt ctttgtttat
3901 gagtattgct gatatgtggc tagctgccac acttttcttg tcctttccat ttacaataaa
3961 tatctttcta tctccacca aattaaagta ctccgcaacc tgttattcca ccagcatcc
4021 cttccctctt caactacaat ttcatgcagc gatcaagaaa tagaatgtac cgactgtttg
4081 ccattggtgg ggcattggaa aagtgggtgg aaagtgcaga gcttagatta taaaggccag
4141 ggtgagagtt cccattgtgg tcagctgaaa tgaatctgac tagcatccat gagcacgaag
4201 gtttgatccc tggcctcaat cagtgggtta aggatctggc gttgctgtcc gtgagttgtg
4261 gtgtagttcg cagacaaggc gtggacttag tgtggctgtg gctgtggcat aggctagtgg
4321 ctacagctct gattcgacct ctgacctggg aatctctata tgctgtgagt gtggccctaa
4381 aatttaaata aaattaaata aaggaccagg gtatattttt ctttggagat aaggtacata
4441 gtcagtatat caaggacagt agacctagga aacggatgct tccctagtc ttgtatgca
4501 ggtggggcat ctgagttggg ggcggctgga gcccttaggg accattaact aaaccgtca
4561 ctctcccaca tctcgggtga ccttgggtatc agtcaggatg cttccctttt gagcctcaaa
4621 atggccttag tctccttccc aaccagacg gccctgtcag ttcattgact tggctaattt
4681 gccagtgtag gcctatgcaa attaaggtag aacgcactcc ttagcgctcg ttgactatc
4741 atcaactttt ctttttagaa aagatattgg tataagcact tcttaaaaaa ccatattcca
4801 ctctgggtgt atttaactta attttccctt ctccttttct tttcccaggA G 3' end of

```

INTRON 3 -----puromycin bpolyA----- beginning of intron 4

```

4938          GTa attatgaaac atgatgaat gatgttgatg aaagtctcct
4981 ctaatctcct agttatcagc caagtcacca gcttgcatta aaagtaggat tcaactgacac
5041 cgtaagaaa gcattccaga gattgcccgt tgtggctcag gggcagcaaa cccaattagg
5101 atccaagagg aggtgggttt gatccctggc cttgctcttt ggcttaagga tccggcattg
5161 ccgtgacctg tgggtgtagt tgcagatgca gctcggatct ggcattgctg tggctgtggc
5221 gtaggctggg ggcttcagct ccagtttgac ccctagcctg ggaacttcca tatccacac
5281 ttgcccctt aaaaatcaaa gaaagaaaga aaatattcta cccttctgt atccctgagc
5341 ccttaaatat cgtcttttaa gtcattagat cttcaagtac cttccagcta attaatatc
5401 ttccttctct ccatgttgcc attgtcctga tttttatacc tctgcagttc tgggtaggct
5461 agagccagaa ataataaggt catgttaaga ccaagatata atattaaatt atttatatga
5521 ccagatatgg aagttacctt gagaactttc agacaggaat tccatgagaa atacaccctg
5581 atttttgcaa tcttaaaata tttgcagagt ttaaaggaa aactcaagtt gttgactttt
5641 gctgcaaaac acactgagtc gctggtgatt catttgtgcc tggctaaact tttgggtgtt
5701 ttgtcttttt ttttttttaa ctctggaaag caaatgaat taaacatttc tgagttttca
5761 aattcatcag tggattcacc ccaaatattt gagctgcttc tttgcttttg gaaactacga
5821 tgccttggag attccagctg gagacgcttc tgacagaaag aaatgtctgc aagcagctac
5881 aaaaatgcat gatggctttg acttaagagg cattgatacc gcttggcctt tctttcaaaa
5941 aggccacctt acaacttggc ctgaaggcat tcccgtcgtg gtgcagcgga aaatgaatct
6001 gactaggaac cccgaggttg tgggttcaat ccctggcctt gctcagtggt ttaaggatcg
6061 ggtgttgaag taagctgtgg tgtagattgc agacgcagct tggatctggt gttgctgtgg
6121 ctttgggtga ggcgggcagc tacagctcca cttggacccc tagtctggga accttttagg
6181 tgtggcccta aaaggaaaaa agacaacaaa caaacaaaaa accaaaaaac aacttggcct
6241 ggagagctat gtcacacca ttgatatttt gatgggtagt gtttttagtag cccctcaagt
6301 tcaggatgat ggcctggatt aacattagaa tgtctcttaa attctacgac ttgatgagcc
6361 agcaggacca ttttggccac ttagaaagga actgcactct caggtccatc agtagaagga
6421 ggattctcta gggagtctc tcttagctca ggggttcaa gaattcagtc ttgtccctac
6481 agcagctcag gtgactgcta tggcttggct ttgatccctg gcccaggaat tctgcatgc

```

24/27

6541 tgcaggtgca gccaaaaaaa aaaaaaaaaa aaggaggagg nggattccct agaataagaa  
6601 gctatcattc ctttggatgc ttcataagatc taaccacttc tggaacagtt attccctctc  
6661 attctgaaga actcatttta agaaaaacaa gacgagctag agagtgaaca aatgggtctac  
6721 aaaccaggcc tttcgaattg aggaaactgt ggtacttcct ctgaagaaaa gatgacagcg  
6781 ttggatgcag agaccctggg gctcccttag gtacttgagg actgaggaga tattctcagt  
6841 ggaggctgga gctaggctgc ctggggctgg tcctgtgcca ccacttcctt cctctgtgac  
6901 tttgggcaag tttccctatc tttaaaaatg gggatgatag tagtacctgc ttcatagggt  
6961 tgttggataa aataagttgt gaataaagca ctaaggggca cgtacttagt aagcgctggc  
7021 tgccatcacc accaccacta tcaccatctg tccggagggc agcataggac aggagatttt  
7081 tggcaaatag aaggaagagt tctaggagtt cccgttgtgg tgcaggggaa atgaatccaa  
7141 ctaggacta ggaggtttcg ggttcaatcc cgcgcctcgc tcagtgggtt aaggatccag  
7201 tgttgccatg agctgtggtg tagattgcag acatggctag gatctggagt tgctatggct  
7261 gtggtgtaag ctggcagctg tagctcggat tctacccta gcctgggaat tccgtagtc  
7321 cacaggtttg gccctacaaa gaaaaaagaa aaagaaaaag aaaaaattct aggggctgaa  
7381 agaataaac agaagagcaa gttcccatg gggttcctga cctgagttga gatgcttggtg  
7441 taggcaacct tcaagctctg aactcttgat tgttttgaat tgcagccaga gttgtacttc  
7501 catattttgg gtacttcaca aaattaaaac acagaagcca aaggcccaga agtgcataatt  
7561 ggtgctggcc tcccataaag aggggtgttt tgcaagtctg ggcacactct ctcttcacag  
7621 taactggagc agattctggc tgctcttcag ggccgtagtc tggcaccag actgcagcca  
7681 catcattctt caatgtgagg aatctatttg aacatctgca aggggtttaa aaggcaggag  
7741 attctttgcc acctgtgtaa ttggtctgag gtgagctgag ggcactaac ttagacagggt  
7801 gggtagcact gtagctaaag aggattacag gagtccctgt tgtggcttag tggtaacaaa  
7861 tccaactagt atccatgagg attcagggtc gatccctggc ctgctcaggt gggtcaggta  
7921 tccggtgttg ctgtggctgt ggtgtaggct ggcagcttca ttttatttac ccctagcctg  
7981 ggaacttcca tgtgctgtag gtaaggccct tgaaaaaaa aaaaaagaga tttcaaaata  
8041 actccatcaa acacatacag ctgtttaaga atgtcatcca ggacagcatt tgggtaaagg  
8101 ctagatgaaa aaaaaaaaaa aaaaacttag aattttattt atttattttt tctttttagg  
8161 gccagacctg tggcctatgg aaatgcctgg gctaggggtg gaatcagagc tgcttacacc  
8221 acagccatag ccacgccaga tccaagcccc gtctgtgacc tacaccacag ctcatggcaa  
8281 aactggatc cttaatccac tgagtgaggc caggaattga acccacattc tcatggatgc  
8341 tagttgggtt cttaagccac tgagccacaa gcttagaatt ttagagggtg aagaaacttt  
8401 aagagctata ataaagtaat gatggtgatg gtgatttga tgttagcggc tactagttat  
8461 tgagtgtttg cttgtgccag gaactccact gttcattccc tctgtttttt aaacagccc  
8521 tggaagggtc gtgttagtcc acatttctag atgaggaata ctgagtttcc acaatattaa  
8581 atgtgaacgt tcaagggtcac attttttagga agatttaggt ccagggtgtg ctgacttggg  
8641 taacctgggt aaccttccct ttagtcaagg gttccattgt tcaggcgatg aaatggagac  
8701 ccagtagggt aaatgactta acagtgaact tatgtccaac ttctaattag aactcagatc  
8761 ttctgattca tcatctgggg ctcttctctg agctgggtgt tgatgccaaa tgctgagagg  
8821 ggtacagtgt gccgtcaagg agaatcccta ccctcaaggg gttatgctgt agatggagca  
8881 ggcagaggta cccatgaaag cccaacaaca caggctagaa ggaggatgtc agagagagag  
8941 agcaaaaggaa cgtgagagtt cagggagggc aagattatgt ttggcttgga gatggatcta  
9001 tgttttgcac ttattttttt gggggggggg tctttttgct acttcttggg ctgctcccga  
9061 ggcataatgga ggttcccagg ctaggggtct aattggagcc gcagccacca gcctacacca  
9121 gagccacagc aacgcaggat ctgagccgcg tctgcaacct tcaccacagc tcacggcaac  
9181 ncgggatcgt taacctactg ttaagcactg cgccacgacg ggaactccct catttagaaa tatttattga  
9241 gtcagattcg ttaagcactg ttaagcactg cgccacgacg ggaactccct catttagaaa tatttattga  
9301 gcacctaact tatgccaggc attgtgctag gttcatacca aagaaggctc agaagagatg  
9361 gcatccgagc tgtgccttga aggatgaata tgtgttaaat gccgtacact tcagggtggg  
9421 tgttgcctgt acctgagggt ttgaaggctt ctgggaaagg aggggtgagat gaggaagagg  
9481 gaggggttac taaaaagatg ggacgagggt gcaaatccaa atctataaat tgatgccctg  
9541 agtgccctgc aggagggtgg ggctcctgag tgctgggtgg cacgggccct tccccctcct  
9601 cttgccctct tcccttcccc ctcttctgag atctgaagtc agattcccca ggttcaaaata  
9661 ctgtttcttc ccttagcagt atgacctgg gcaaaataat ttattgcctc tgtccctctg  
9721 aggaggaata gaacctcctt cattgactgt tattagaatt taatgagcta atacatgtca  
9781 gttgcttaga aagggtccca gccaaactatt agctattatg aatattatca gatcaataga  
9841 cagatttaga aacaagggac ttttagagctg ggtccatggg tactgagctt agaggggaaa  
9901 ccataggtgg taggaaggca tgtatttcat tcctaccagg agatgtggac tcccagctgg

25/27

9961 ggcagaaggc agagggagga gatcggggct ttggcagaat ctcaaacaaa tattagtggg  
10021 tagtgggtttt ttgtttctgt ttttaagagat gagggcaggc gtttccgatg tggcgagtg  
10081 gaaacgaatc tgactagtat ccatgaggat gcaggttcca tccctggact cactcagtga  
10141 gttaaggatc cggcattgcc gtgagttgtg gtgtnggtca cagacacagc tcagatctgg  
10201 cattgctgtg gctgtggtgt aggctagcag ctgtagctcc aattcaaccc ctagecctggg  
10261 aacttccaca tgccgcagggt gcaaccccaa aagataaatg aataaataaa taaatatgcg  
10321 accttccttt cttggggcct ttgcatgttt ttctctctgt taggcacact cttgctaate  
10381 cctcttcact gggcctccta agtatecttc agaactcagc taaaacatca tccccctccc  
10441 tggggagcct tcgagggtctt cctgttaagt gctcctatgc tttcttggag ttttgaagtc  
10501 ctataatgat gtgtttatca aaatagggtc caccctccct gccagcttct ctacaccaca  
10561 gacacatggt gtctgtttca gtcaacactg tatgtctggc acttgacatg taacgcagtc  
10621 tcagcaggta tttgttgaat gaatggaggc ggtctgctag agtcgtcata tttttactga  
10681 tcccgctcttg taggatgggtc tctactgctt tgttagctta agaagtacct tttttttttt  
10741 tttttttaat ggccacaccc atggcatata gaaattccac gaaggaagga agaaagaaag  
10801 aaagaaagaa ggaaattcct gggtcaggga ttgaatccaa gccacagggtg caacctgagc  
10861 tgcagttgtg gcaacaccac atcctttaac ccactgtgct gggccaggga tcatacctgt  
10921 gcatgtacag cgacccaagc cacggcagtc agattctttt tcttctcttc tttctttctt  
10981 ttcttttctt tttttttttt tttttttttt ggctttttgc cttttctagg tgcggcata  
11041 ggaggttccc aggctagggtg tcgaatcaga gctgtagacg ccggcctaaa ccacagccac  
11101 agcaacacag gatccaagcc ttgtctgtga cctacaccac agctcacggc aacgctggat  
11161 ccttaacccg ctgagcgagg ccagggattg aaccgcgaac ctcatgggtc ttagttggat  
11221 tcgttaacca ctgagccatg atgggaactc ctgcagtcag attcttaacc cactatgcca  
11281 cagcaggaac tctagaagt gccctttgag gctactctgt agacagctct gagccagcga  
11341 ggcaagacct gtttttctgg aggaagataa atcctgggtg agggatgggt gggctgtggt  
11401 cttcctggga cccatctctg gagcctctct ccctcagcaa agccaccttg gacaataaga  
11461 gctgccatct attttttttt ctttaaacta agatttgata ttttccagag acctccctc  
11521 ccaccgtttg atctgagtaa ttctgaaatg acgagagtcc cgtgatatca ttttttcgat  
11581 ctcgaagggtg gaaacctggg agtagccaca acccaggctc tcagctcagc ctagggtttc  
11641 aatgataatg attgcaaaat agcttttctc tgcatccaa gtaacatgat atgtttttat  
11701 ttccatttgc ttttag //

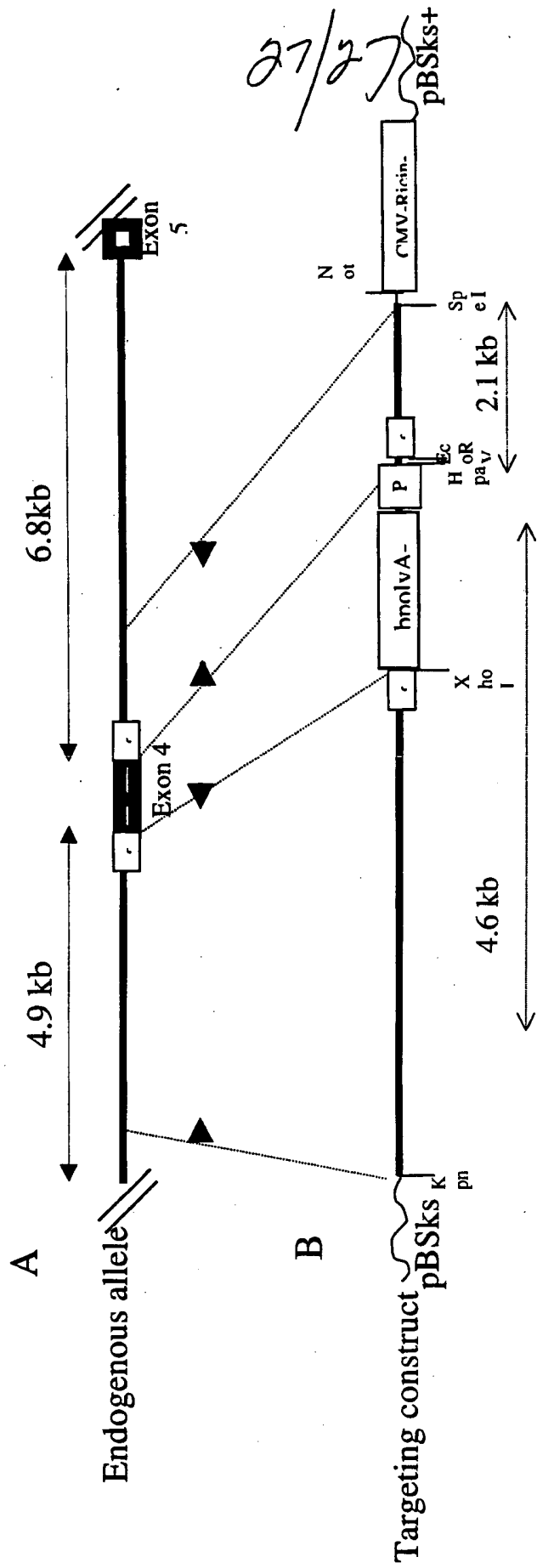
26/27

**Figure 8**

1 gaattcgcgga atccggagtg taacatATGA TATTTCCCAA ACAATACCCA ATCATAAACT  
61 TTACCACTGC AGGTGCCACT GTGCAAAGCT ACACAAACTT CATCAGAGCT GTTCGCGGCC  
121 GCTTAACAAC TGGAGCTGAT GTGAGACATG AAATACCAGT GTTGCCAAAC AGAGTTGGTT  
181 TGCCTATAAA CCAACGGTTC ATACTAGTTG AACTCTCAA TCATGCAGAG CTTTCTGTTA  
241 CATTAGCTCT AGATGTCACC AATGCGTATG TGGTCGGCTA CCGTGCTGGA AATAGCGCAT  
301 ATTTCTTTCA TCCTGACAAT CAGGAAGACG CCGAAGCAAT CACTCATCTT TTCCTGACG  
361 TCCAAAATCG ATATACATTC GCCTTTGGTG GTAATTACGA TAGACTCGAG CAACTTGCTG  
421 GTAATCTGAG AGAAAATATC GAGTTGGGAA ACGGTCCGCT AGAGGAGGCT ATCTCAGCGC  
481 TTTACTATTA CAGTACTGGT GGTACCCAGC TTCCAACCTT GGCTCGTTCC TTCATAATAT  
541 GCATCCAAAT GATTTTCAGAA GCAGCAAGAT TCCAATATAT TGAGGGAGAA ATGCGCACGC  
601 GTATTAGGTA CAACCGGAGA TCTGCACCAG ATCCTAGCGT GATCACACTT GAGAATAGTT  
661 GGGGGAGACT GTCGACTGCA ATTCAAGAGT CTAACCAAGG AGCCTTTGCT AGTCCAATTC  
721 AACTGCAAAG ACGTAATGGT TCGAAATTCA GTGTGTACGA TGTGAGTATA TTAATCCCTA  
781 TCATAGCTCT CATGGTGTAT AGATGCGCGC CTCCACCGAG CTCACAGTTT TAGTAAGgat  
841 ccaagctt

**Figure 9**

**Gal $\alpha$ 1-3Galactosyl Transferase Collision Intron Targeting Construct**



▶ : PCR primer positions

SA: splice acceptor, AG dinucleotide

SD: splice donor, GT dinucleotide

PGK-Puro-bpA: PGK promoter, puromycin gene coding sequence

and bovine growth hormone poly A signal sequence

pBS Ks<sup>+</sup>: Bluescript cloning vector

CMV-Ricin-SV40pA: CMV promoter, Ricin gene and SV40 poly A

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
5 April 2001 (05.04.2001)

PCT

(10) International Publication Number  
**WO 01/23541 A3**

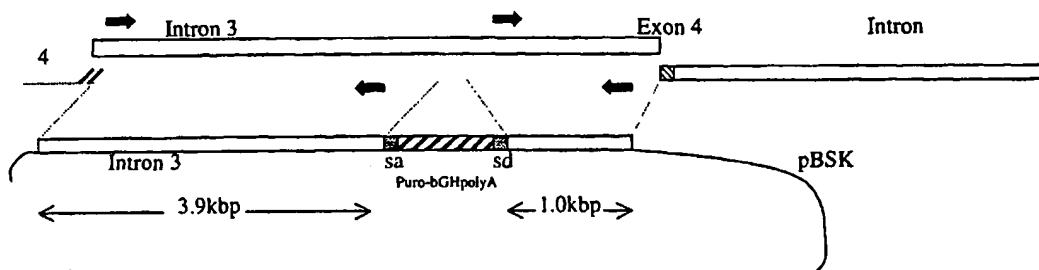
- (51) International Patent Classification<sup>7</sup>: **C12N 15/54**, 15/90, 15/63, 15/67, 15/85, 9/10, 5/10, 1/21, A01K 67/027
- (21) International Application Number: PCT/US00/27065
- (22) International Filing Date: 2 October 2000 (02.10.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/156,953 30 September 1999 (30.09.1999) US
- (71) Applicant (for all designated States except US): **ALEX-ION PHARMACEUTICALS, INC.** [US/US]; Suite 360, 25 Science Park, New Haven, CT 06511 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **FODOR, William, L.** [US/US]; 236 Wildcat Road, Madison, CT 06443 (US). **RAMSOONDAR, Jagdeee, J.** [TT/US]; 46 Madison Street, Hamilton, NY 13346 (US).
- (74) Agents: **WIXON, Henry, N. et al.**; Hale and Dorr LLP, 1455 Pennsylvania Avenue, N.W., Washington, DC 20004 (US).
- (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— with international search report
- (88) Date of publication of the international search report:  
17 January 2002

[Continued on next page]

(54) Title: COMPOSITIONS AND METHODS FOR ALTERING GENE EXPRESSION

Promoter trap Construct: Insertion of Engineered exon into the intron of a gene.

Endogenous alpha-Gal transferase gene diagram



PCR Primers →

sa = splice acceptor

sd = splice donor

pBSK = pBluescript cloning vector

Puro-bGHpolyA = puromycin resistance with the bovine Growth hormone gene poly A acceptor signal sequence

(57) Abstract: Disclosed are novel compositions and methods useful for modulating expression of a target gene in a cell by insertion of exogenous DNA sequence into the target gene. The compositions and methods of the invention are useful for generation of knockout animals including mammals.



*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



# INTERNATIONAL SEARCH REPORT

International Application No

PL/US 00/27065

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/54 C12N15/90 C12N15/63 C12N15/67 C12N15/85  
C12N9/10 C12N5/10 C12N1/21 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, BIOSIS, PAJ, EMBL

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KARAPLIS ET AL: "Gene targeting" PRINCIPLES OF BONE BIOLOGY, XX, XX, 1996, pages 1189-1201, XP002090123  figures 3,4  ---	1-9, 24-38, 47,51, 62,65,68
X	US 5 916 810 A (JARVIK JONATHAN W) 29 June 1999 (1999-06-29)  the whole document  ---  -/--	1-9, 24-38, 47,51, 62,65,68

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

11 June 2001

Date of mailing of the international search report

16 07. 2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Galli, I

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/27065

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 95 28412 A (GUSTAFSSON KENTH T ;INST OF CHILD HEALTH (GB); BAETSCHER MANFRED W) 26 October 1995 (1995-10-26)	1-13, 19-23, 25-41, 43-53, 55,62-72
A	examples 6,7	
P,X	WO 99 61604 A (JAPAN SCIENCE & TECH CORP ;LUKACSOVICH TAMAS (JP); AWANO WAKAE (JP) 2 December 1999 (1999-12-02)	1-9, 24-38, 47,51, 62,65,68
E	abstract; figure 5	
A	WO 01 30992 A (KOIKE CHIIRO ;UNIV PITTSBURGH (US)) 3 May 2001 (2001-05-03) compare nt 10070-12219 of seq. 1 of present application with nt 493-2651 of seq. 10.	56,57
A	HOOPES C.W. & PLATT J.L.: "A molecular epidemiological probe for pig microchimerism" TRANSPLANTATION (BALTIMORE), vol. 64, no. 2, 1997, pages 347-350, XP000982842 figure 1	13,20, 41,44,46
A	US 5 627 058 A (RULEY H EARL ET AL) 6 May 1997 (1997-05-06) abstract	24
A	US 5 464 764 A (THOMAS KIRK R ET AL) 7 November 1995 (1995-11-07) table 1	72

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 00/27065

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
  
Although claims 68-70 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 54  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

As a result of the prior review under R. 40.2(e) PCT,  
all additional fees are to be refunded.

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☒ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: partly 1-13, 19-41, 43-55, 62-72

A method of modulating the expression of a eukaryotic gene in a cell, comprising (i) transfecting a cell with a nucleic acid construct, said construct comprising a first construct sequence homologous to a first gene sequence, a sequence encoding a selectable marker, and a second construct sequence homologous to a second gene sequence, wherein said first and second gene sequences independently comprise at least a portion of one or more intron regions of said eukaryotic gene, and (ii) integrating said selectable marker into said eukaryotic gene, wherein expression of said selectable marker results in modulation of said eukaryotic gene in said cell. Corresponding nucleic acids, vectors, recombinant host cells, transgenic animals and related methods and applications.

Said method, materials and applications, wherein said intron region is intron 3 of porcine the alpha-1,3-galactosyl-transferase gene.

2. Claims: 14, 56 and partly 1-13, 19-42, 43-55, 62-72

Idem as subject-matter 1, but limited to intron 4.

In addition, a nucleotide sequence of intron 4 having nucleotides 4938-11716 of Fig. 1.

3. Claims: 15, 57 and partly 1-13, 19-42, 43-55, 62-72

Idem as subject-matter 1, but limited to intron 5.

In addition, a nucleotide sequence of intron 5 having nucleotides 11753-13748 of Fig. 1.

4. Claims: 16, 58 and partly 1-13, 19-42, 43-55, 62-72

Idem as subject-matter 1, but limited to intron 6.

In addition, a nucleotide sequence of intron 6 having nucleotides 13810-14358 of Fig. 1.

5. Claims: 17, 59 and partly 1-13, 19-42, 43-55, 62-72

Idem as subject-matter 1, but limited to intron 7.

In addition, a nucleotide sequence of intron 7 having nucleotides 14463-21627 of Fig. 1.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

6. Claims: 18, 60 and partly 1-13, 19-42, 43-55, 62-72

Idem as subject-matter 1, but limited to intron 8.

In addition, a nucleotide sequence of intron 8 having nucleotides 21766-27048 of Fig. 1.

7. Claims: partly 1-13, 19-41, 43-55, 62-72

Idem as subject matter 1, but limited to intron 9.

8. Claim : 61

A lambda phage clone derived from porcine genomic library comprising at least a portion of the Gal alpha-1,3-gaactosyltransferase gene, wherein the lambda phage is selected from the group consisting of pgGT, lambda1, lambda2, lambda4-1 and lambda8-2.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 54

Claims 10 and 39 relate to nucleic acids and methods involving a gene from a panel of many possible genes to be targeted. The description provides support only for Gal alpha-1,3-galactosyltransferase. Pursuant to Art. 5 PCT the search has been restricted to this latter gene.

Claim 54 is missing.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

F /US 00/27065

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5916810 A	29-06-1999	US 5652128 A	29-07-1997
		EP 0656954 A	14-06-1995
		WO 9416092 A	21-07-1994
		US 6096717 A	01-08-2000
		US 6080541 A	27-06-2000
-----			
WO 9528412 A	26-10-1995	AU 1850599 A	29-04-1999
		AU 2233295 A	10-11-1995
		CA 2187802 A	26-10-1995
		EP 0755402 A	29-01-1997
		JP 10504442 T	06-05-1998
		US 6153428 A	28-11-2000
-----			
WO 9961604 A	02-12-1999	JP 11332564 A	07-12-1999
		EP 1078050 A	28-02-2001
-----			
WO 0130992 A	03-05-2001	NONE	
-----			
US 5627058 A	06-05-1997	US 5364783 A	15-11-1994
-----			
US 5464764 A	07-11-1995	US 5487992 A	30-01-1996
		US 5627059 A	06-05-1997
		US 5631153 A	20-05-1997
		US 6204061 B	20-03-2001
-----			

**THIS PAGE BLANK (USPTO)**